Collective Information Processing and Criticality, Evolution and Limited Attention.

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

D o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M. Sc. Pascal Peter Klamser

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Bernhard Grimm

Gutachter:

1. Dr. Pawel Romanczuk
2. Prof. Jens Krause
3. Prof. Iain Couzin

Tag der Einreichung: 21. Dezember 2020
Tag der mündlichen Prüfung: 11. März 2021

Abstract

In this work, I approach coordinated motion of individual agents and its information spread from the biological and statistical physics perspective. In the first part of this work, I will focus on a prominent idea originated from statistical physics, the self-organization to criticality, and combine it with the fundamental biological principle, the evolution. To be precise, the analyzed critical point is the phase transition from an ordered to a disordered phase. In models of statistical physics with an analog phase transition the responsiveness of the system to external signals is largest at this critical point. I investigate the idea of optimal responsiveness at criticality in a biological setting by simulating a cohesive swarm of interacting agents which needs to avoid a pursuing predator. The latter represents the external signal. From an outside perspective it is appealing to assume that the collective self-tunes itself to the critical point to best respond to the predator. I test this criticality hypothesis (also prominent in other interdisciplinary fields as neuroscience) for group living animals by applying evolution as the tuning mechanism.

It turns out that (i) the best group avoidance is at criticality but (ii) not due to an enhanced response but because of structural changes which are fundamentally linked to the phase transition. Most importantly (iii) this group optimum is not an evolutionary stable state, in fact (iv) at the phase transition evolution drives the system away from the order disorder transition, i.e. it is an evolutionary accelerator. The reason for this accelerated evolution is (v) a maximal spatial self-sorting of individuals (again fundamentally linked with the transition) and consequently strong spatial selection.

The second part of this work models experimentally observed differences in collective behavior of fish groups subject to multiple generation of different types of size-dependent selection. The real world analog to this experimental evolution is recreational fishery (small fish are released, large are consumed) and commercial fishing with large net widths (small/young individuals can escape). In both cases the fish population is meant to be preserved, but does this size selective pressure induces evolutionary behavioral changes? In addition, how do those possibly behavioral altered individual perform collectively in the natural environment under natural predation? The fish of the large harvested selection line are less cohesive and in the same time less risk taking. However, cohesion is considered as a main defense mechanism against natural predators. We hypothesis that both findings are in agreement with a mechanistic explanation based on an attention trade-off between social and environmental information. I numerically justify the hypothesis, by fitting a mechanistic model to the different selection line. Furthermore, with the model representations of the selection we predict the performance in a natural predation and commercial fishing context. It turns out that those measures (only harvest the large individuals) which are meant to preserve the population, might degrade the defense mechanism against natural predators and in the same time allow the shoals to better avoid fishing nets.

The last part of the thesis is an attempt to quantify the collective information processing in the field. The study system is a fish species adapted to sulfidic water conditions with a collective escape behavior from aerial predators which manifests in repeated collective escape dives. Methods are introduced to quantify this collective behavior at different scales. Since the fish only measure about 2 centimeter but the collective wave spreads across meters in dense shoals at the surface, the investigation at different scales is crucial. On the large (macro) scale, I find differences in the wave propagation speed depending on the direction relative to the shoal orientation. On the small (meso) scale, I find that wave speed increases weakly with polarization, is fastest at an optimal density and decreases with increasing difference in the angle between the wave propagation direction and the local shoal orientation.

Zusammenfassung

In dieser Arbeit nähere ich mich der koordinierten Bewegung einzelner Individuen und ihrer Informationsverbreitung aus der Perspektive der biologischen und statistischen Physik an. Im ersten Teil dieser Arbeit werde ich mich auf eine Idee konzentrieren, die aus der statistischen Physik stammt, nämlich die Selbstorganisation zur Kritikalität, und sie mit dem grundlegenden biologischen Prinzip, der Evolution, verbinden. Genauer gesagt ist der analysierte kritische Punkt der Phasenübergang von einer geordneten zu einer ungeordneten Phase. In Modellen der statistischen Physik mit einem analogen Phasenübergang ist die Reaktionsfähigkeit des Systems auf äußere Signale an diesem kritischen Punkt am größten. Ich bewege dieses Konzept in einen biologischen Rahmen, indem ich einen kohäsiven Schwarm interagierender Agenten simuliere, der einem verfolgenden Jäger ausweichen muss. Der Jäger repräsentiert hierbei das externe Signal. Von außen betrachtet reißt die Annahme, dass sich das Kollektiv selbst auf den kritischen Punkt einstellt, um am besten auf das Raubtier zu reagieren. Ich teste diese Kritikalitätshypothese (die auch in anderen interdisziplinären Bereichen wie der Neurowissenschaft populär ist) für gruppenlebende Tiere, indem ich die Evolution als Organisations-Mechanismus anwende.

Es stellt sich heraus, dass (i) die Gruppe den Jäger am besten am kritischen Punkt ausweicht, aber (ii) nicht aufgrund einer verstärkten Reaktion, sondern aufgrund struktureller Veränderungen, die grundlegend mit dem Phasenübergang zusammenhängen. Am wichtigsten ist, dass (iii) diese Gruppenoptimum kein evolutionär stabiler Zustand ist, sondern (iv) der Phasenübergang als evolutionärer Beschleuniger agiert, der das System vom Ordnungsstörungsübergang wegdrängt. Der Grund für diese beschleunigte Evolution ist (v) eine maximale räumliche Selbstsortierung der Individuen (wiederum grundlegend mit dem Übergang verbunden) und als Folge davon eine starke räumliche Selektion.

Der zweite Teil dieser Arbeit ist motiviert durch Unterschiede im emergenten kollektiven Verhalten zwischen Fischgruppen, deren Vorfahren über mehrere Generationen hinweg nach unterschiedlichen Größenkriterien ausgewählt wurden. Das reale Analogon zu dieser experimentellen Evolution ist die Freizeitfischerei (kleine Fische werden zurückgesetzt, große werden verzehrt) und die kommerzielle Fischerei mit großen Maschenbreiten der Fischernetze (kleine/junge Individuen können entkommen). In beiden Fällen soll die Fischpopulation geschont werden, aber induziert dieser größenselektive Druck evolutionäre Verhaltensänderungen? Und wie verhalten sich diese möglicherweise verhaltensveränderten Individuen kollektiv in der natürlichen Umgebung unter natürlicher Prädation? Die Fische von deren Vorfahren nur die kleinen überlebten sind weniger kohäsiv und gleichzeitig weniger risikofreudig. Aber der Zusammenhalt (die Kohäsion) wird als ein Hauptabwehrmechanismus gegen natürliche Räuber angesehen. Wir stellen die Hypothese auf, dass beide Befunde mit einer mechanistischen Erklärung übereinstimmen, die auf einem Aufmerksamkeitskompromiss zwischen Sozial- und Umweltinformation beruht. Ich rechtfertige die Hypothese numerisch, indem ich ein mechanistisches Modell an die verschiedenen Selektionslinien anpasse. Darüber hinaus sagen wir mit den Modelldarstellungen der Selektionslinien die Überlebensrate in einem natürlichen Raubtier- und kommerziellen Fischereikontext voraus. Es stellt sich heraus, dass diejenigen Maßnahmen (nur die großen Individuen fischen), die die Population erhalten sollen, den Abwehrmechanismus gegen natürliche Raubtiere herabsetzen und gleichzeitig den Schwärmen erlauben könnten, Fischernetze besser zu vermeiden.

Der letzte Teil der Arbeit ist ein Versuch, die kollektive Informationsverarbeitung im Feld zu quantifizieren. Das Studiensystem ist eine an sulfidische Wasserbedingungen angepasste Fischart mit einem kollektiven Fluchtverhalten vor Raubtieren aus der Luft, das sich in wiederholten kollektiven Fluchttauchgängen manifestiert. Es werden Methoden vorgestellt, um dieses kollektive Verhalten auf verschiedenen Skalen zu quantifizieren. Da die Fische nur etwa 2 Zentimeter messen, die kollektive Welle sich aber in dichten Schwärmen an der Oberfläche über Meter ausbreitet, ist die Untersuchung auf verschiedenen Skalen entscheidend. Auf der großen (Makro-)Skala finde ich Unterschiede in der Wellenausbreitungsgeschwindigkeit in Abhängigkeit von der Richtung relativ zur Orientierung der Schwärme. Auf der kleinen (Meso-)Skala stelle ich fest, dass die Wellengeschwindigkeit mit der Polarisation schwach zunimmt, bei optimaler Dichte am größten ist und auch, wie auf der Makro-Skala beobachtet, vom Winkelunterschied zwischen der Wellenaus-

breitungsrichtung und der lokalen Orientierung des Schwarms abhängt.

Acknowledgments

First, I would like to thank Pawel Romanczuk, who created an unique working environment, whose advice has always helped me and who has supported me and inspired me in every possible way. Also for bringing the scientific world to us so vividly through a variety of conferences and collaborations. I owe many unique encounters, both human and scientific, to the CoCCoN network he initiated. Especially in these times of isolation it is a pleasure to think back.

I would like to thank the people who shaped my daily work at the lab. It was a great pleasure for me to work with Winnie Poel, explore the research field of collective behavior and pursued the path to dissertation together. I owe Yinong Zhao many enthusiastic conversations about related scientific fields. Haider Klenz has humorously lit up the office and was a highly valued sitting neighbor. I was fortunate to explore the role of a supervisor with the talented interns Conor McBride, Alexandra (Olenka) Jain and Noam Miller. I would like to thank Luis Alberto Gomez Nava, Fritz Francisco and Mohsen Raoufi for enlightening conversation.

During my adventures and research in Mexico and the fieldwork there, I was able to enjoy the company of Carolina Doran, David Bierbach, Juliane Lukas, Pawel Romanczuk, Jens Krause, Haider Klenz, Marie Habedank, David Luis and Felix Auer. Those were times with people I would not want to miss.

I also thank Valerio Sbragaglia, with whom I joyfull explored zebrafish trajectories together with Robert Arlinghaus and Pawel and thus experienced my first interdisciplinary cooperation.

My everyday working environment was characterized by people and institutions that made scientific and human interaction pleasant, eventful and beneficial. The broad spectrum of the seminar at the Institute of Theoretical Biology(ITB) was an enrichment and challenge and its members were incredibly friendly. Karin Winkelhöfer, Margret Franke and Jana Lahmer were a great help in all administrative matters.

I deeply thank my family and friends for all their support they gave me. Especially my daughter, Gwendolin, to whom I owe so much joy and with whom I was able to discover the world anew. Especially through the eyes of a girl, through whom I have learned to appreciate the great women of our time in order to relativize and soften the role models of our society.

Contents

List of Figures

List of Tables

1 Introduction

The relationship between predator and prey has been studied with mathematical models for at least a century. One of the best known models that conceptualize the interdependence of predator and prey populations is the Lotka-Volterra model. It was independently published in 1925 by Alfred J. Lotka and in 1926 by Vito Volterra (Kingsland, 1985). Interestingly, Vito Volterra was inspired by the data collected by the marine biologist Umberto D'Ancona, who was surprised by the statistics of catches in the Adriatic Sea. He was amazed that more predatory fish were caught although their prey fish became rarer (Kingsland, 1985).

In the last decades we could observe a similar interplay between data and theory. The data from a century ago allowed the rough estimation of the population size and therefore inspired mathematical models of population dynamics. The data from the recent past allows the reconstruction of individual trajectories of individuals in swarms (Buhl et al., 2006; Cavagna et al., 2010; Francisco et al., 2020; Graving et al., 2019; Handegard et al., 2012; Strandburg-Peshkin et al., 2013). This precise nature of the data inspired models that depict the detailed movement or decision behavior of schooling fish (e.g. Calovi et al., 2018; Gautrais et al., 2012; Sosna et al., 2019).

However, especially in the field of collective behavior the numerical simulation of self-propelled agents with social interaction even pioneered the biological data collection (Couzin et al., 2002; Vicsek et al., 1995). For example, Couzin et al. (2002) showed that a collective memory can exists even if the individuals have no memory at all.

The numerical exploration of models made it also possible to answer questions beyond the lifetime of individuals. Why do animals group together? There existed theoretical explanations which highlighted the benefits of group living (Hamilton, 1971; Krause and Ruxton, 2002).

However, to our knowledge no mechanism was explicitly tested to explain the emergence of cohesive and coordinated movements in agent-based systems, until more than a decade ago (Jamie Wood, 2010; Olson et al., 2013, 2016; Wood and Ackland, 2007). This was done by simulating the interaction of predator and prey and on top of that also the evolution itself. In fact those studies identified the predator and the confusion effect as the driving mechanism for cohesive groups (Olson et al., 2013, 2016).

A closely related science branch to collective animal behavior is statistical physics. In both system a large amount of identical or similar agents interact with each other. The similarities inspired the adaptation of concepts from statistical physics to collective behavior (Mora and Bialek, 2011). For example, Cavagna et al. (2010) could show that the fluctuations in the velocity of startling flocks are scale free. Scale free behavior is also typical for statistical physics system at the transition from an ordered phase to a disordered state. Just like in a magnet that has no net magnetization if the temperature is too high but becomes magnetic below the Curie temperature. Scale free correlations are often associated with a maximal information transfer between individuals.

- **Q1:** So why should collectives not be in a state where the information transfer is maximal?
- **Q2:** Is evolution driving group living animals to phase transitions where this property is predicted from statistical physics?

Both of these questions are rarely studied in this specific framing. In contrast, a modified combination of them "Is a collective responding better to signals at a phase transition?" was studied extensively (e.g. Calovi et al., 2015; Vanni et al., 2011). It is important to explore not only reasons why a collective should operate at a phase transition, but also possible counterarguments. The second question stresses the fact that we are talking about biological systems in which evolution is possibly the main driver for adaptation. The first part of my thesis will explore exactly these questions. I will model a collective, an attacking predator and evolution to gain insights.

Group living animals nowadays do not only risk ending up as prey of their natural predators. This was not even the case when Volterra published his model, since he was inspired by the fact

that despite the reduced fishing pressure (during World War I) the prey fish population did not increase but the predator population did. Thus, already back then the anthropogenic pressure (fishing) had a major impact on populations. We even know that fishing can change the onset of fish migration (Quinn et al., 2007). In the Anthropocene, where we are discussing how humanity needs to reduce emission to prevent a "Hothouse Earth" (Steffen et al., 2018) we should also consider how anthropogenic fishing pressure can affect the collective behavior of fish.

However, it is very difficult to estimate the impact of fishing on collective behavior in the wild. An alternative is to have an experimental evolution of small groups and extrapolate the results to larger groups. Models have shown that emergent collective behavior can be caused by very different microscopic mechanism (Bastien and Romanczuk, 2020; Jhawar et al., 2020; Romanczuk et al., 2009). Thus, it is important to narrow down the mechanism to be able to extrapolate to larger fish shoals. In this context I aim to answer in the second part of my thesis:

- **Q3:** Which behavioral mechanism is altered by strong fishing pressure?
- **Q4:** Can these changes in behavior influence the survival against natural predators and fishing scenarios?

It is difficult to monitor large groups of animals in the wild. Although in general recent methodological advances improved the situation, for example through video recordings using drones, this is typically not an option for tracking of fish, which do not remain close to the water surface. Also with underwater recordings only parts of the group are captured since the shoals are rarely flat and the three dimensional structure introduces occlusion. If we are interested in responses to predators, it becomes even more difficult. First, data can only be collected if the predator attacks (rare events). Second, the shoal attempt to avoid the predator, inducing a continuous reallocation of the whole system.

All of these challenges do not apply to high density schools of sulfur mollies (*Poecilia Sulphuraria*), a species of fish adapted to sulfidic water. They swim permanently at the water surface and are frequently attacked at the same locations by predators. Most fish respond to predator with a fast acceleration (startle) and in the laboratory it could already be shown that startle responses spread socially, similar to a contagion process (Rosenthal et al., 2015; Sosna et al., 2019).

In this special system this contagious escape response can be quantitatively studied on large scales and the following questions are meant to be answered in the part three, the last part of my thesis:

- **Q5:** Does the fish group responds differently to experimental manipulation compared to natural predators?
- **Q6:** How does the information transfer depend on the collective state and structure?

I will address in each thesis part two of the six listed research questions. The outline of my thesis is as follows.

The first part is purely theoretic and focuses on the evolution under predatory pressure. More specifically, I will test if the collective evolves close to the point where its behavior changes from being ordered to disordered. Therefore, the part is strongly linked with concepts from statistical physics, as phase transitions. I will explore the interplay of group dynamic and structure and the difference between a group and an individual optimum. A topological agent based model of a prey collective is used and the attacking predator is also simulated in an spatially explicit manner.

The second part addresses modeling of zebra fish shoals subject to multiple generation of artificial size-dependent selection pressure. These selection lines are meant to mimic strong recreational fishery (small fish are released) or industrial fishery with nets which allow small fish to escape. I will fit a mechanistic model to the experimentally observed swimming behavior of the groups of the respective selection lines. The model assumes that individuals can not pay attention to social and environmental information in the same time. We hypothesize that this cognitive limitation can explain unexpected differences between the selection lines. Cohesion is one of the differences and is considered as a protection mechanism against predators. Therefore, it is interesting to predict how

the groups would perform in the presence of a predator or in a fishery context. Since the groups are meant to mimic a population under strong fishery pressure, I investigate how the adaptation to fishery impacts the survival against natural predators.

The final part deals with a fascinating biological system in the sulfidic springs of Mexico. There, dense fish schools swim at the surface and can be easily recorded in the field. They also show collective escape behavior in form of escape dives which can propagate through the whole swarm, involving tenth of thousands of individuals. I will introduce methods to quantify these escape behaviors and analyze how the propagation depends on the shoal structure. The specific challenge of the system is that the individuals measure only up to three centimeters but the escape wave can propagate tenth of meters. Therefore, the system is analyzed on different scales with the attempt to answer Q5 and Q6.

Part I

The criticality hypothesis in an evolving collective

The critical point, the region in parameter space where a system changes from one phase to another, is an intensely studied subject in statistical physics (Christensen and Moloney, 2005). The idea that the critical point could be an evolutionary stable state for living systems was inspired by studies of Cellular Automata in the 80s (Bak et al., 1989; Packard, 1988) and 90s (Bak and Sneppen, 1993; Langton, 1990). Conway's "Game of Life", a popular realization of those discrete rule based dynamic models, produces scale-free avalanche distributions and therefore is considered to be close to criticality (Bak et al., 1989). Criticality refers to a transition region between an ordered and disordered/chaotic state and the general reasoning for its importance is appealing: at the edge-of-chaos the system can store information (unlike in the disordered phase) and is still sensible to external signals (unlike in the ordered phase). This combination enables collective computation. Since phase transitions are a general feature of systems with many coupled units the criticality hypothesis spread from statistical physics to biological systems and researchers accumulated evidence in gene regulatory networks (Hidalgo et al., 2014; Roli et al., 2018), neural activity (Beggs and Timme, 2012; Mora and Bialek, 2011; Muñoz, 2018), collective transport in social insects (Feinerman et al., 2018; Gelblum et al., 2015) and collective behavior of cells (Muñoz, 2018), macaques (Daniels et al., 2017), midges (Attanasi et al., 2014b) and starlings (Bialek et al., 2014).

Yet, often the argument of being near-critical is based on a power-law distribution (Bak et al., 1989; Beggs and Timme, 2012). This is indeed a strong indicator because it connects via scale-free correlations (Attanasi et al., 2014b; Bialek et al., 2014) to the phase transitions, but it is not a sufficient condition for being critical (Khaluf et al., 2017). The neuroscience community showed other supporting features (Beggs and Timme, 2012; Haldeman and Beggs, 2005) and for gene regulatory networks there even exists an information theoretic framework, supported by simulations, which labels the disorder-order-transition as an evolutionary attractor under a heterogeneous environment (Hidalgo et al., 2014). Especially this mechanism, the self-organization, is pivotal and is maybe the argument that separates weakly related group living animals from other biological systems. If the gene regulatory network of one individual evolves to criticality because it is more adaptive compared to other non-critical individuals, it is evolutionary on a safe ground. However, even if weakly- or unrelated group-living animals perform collectively better at criticality, only group-selection (Wilson, 1975) would be a likely self-tuning mechanism to criticality, but not necessarily evolution on an individual level (Brush et al., 2016; Torney et al., 2015). Therefore, on the one hand it makes sense that ant-colonies evolve to a phase transition to enhance the cooperative food-transport (Gelblum et al., 2015) because they are strongly related. But on the other hand, it is at least questionable whether this is also true for midge swarms, starling flocks or fish schools, i.e. for unrelated group living animals.

Still, if between-group selection is strong, the unique properties at the phase transition, as a maximal responsiveness to external perturbations (Christensen and Moloney, 2005), might grant

group-fitness benefits. For example, agent-based simulations have revealed that the collective state is strongest influenced by a single individual differing in behavior at the transition (Calovi et al., 2015) or that the synchronization with an external stimuli is maximal there (Vanni et al., 2011). However, such potential benefits have so far not been investigated in spatially-explicit evolutionary models. A characteristic that distinguishes models of collective behavior from classical physics models (Vanni et al., 2011) is the fluid neighborhood. It results in self-sorting of individuals according to their individual parameters (Couzin et al., 2002; Hemelrijk and Kunz, 2005; Jamie Wood, 2010). This is a property with possibly evolutionary consequences because predators might attack certain swarm-regions more frequent (Bumann et al., 1997; Krause, 1994).

There exist numerous studies that investigate predator prey interaction via agent-based simulations. The studies can be clustered with focus on the performance of the prey group (Chicoli and Paley, 2016; Jamie Wood, 2010; Mateo et al., 2017; Olson et al., 2016; Vanni et al., 2011; Wood and Ackland, 2007) or a focus on the predator coordination or catch techniques (Angelani, 2012; Demšar et al., 2015; Janosov et al., 2017) with rare exceptions of a balanced focus (Olson et al., 2013, 2016). Some of those studies that only looked at the group performance but not on evolution of prey avoidance, identified phase transitions as the optimal region of response (Mateo et al., 2017; Vanni et al., 2011). Evolutionary studies found that predatory pressure promotes the evolution of two distinct phases(Jamie Wood, 2010; Wood and Ackland, 2007), interestingly only the one with strong alignment was able to reduce the predator's success (Wood and Ackland, 2007). In a rudimentary agent-based model that does not assume any social forces but where an initially random Markov network evolves, also for the predator, the confusion effect was revealed to be sufficient for the evolution of cohesive collectives (Olson et al., 2013, 2016). van der Post et al. (2015) showed emergence of grouping under predation pressure and interestingly that vigilance decreases with the evolution to larger groups.

Evolution of interacting agents has also been investigated in contexts different from predation. Foraging and predation was explored in Brush et al. (2016); van der Post et al. (2015), where in the latter both agents evolved in both contexts separately and revealed a context dependent evolution not necessarily coinciding with the group optimum (defined by correlation length and robustness). Monk et al. (2018) have shown that foraging in an environment with low exploitation potential and high exploration difficulty a collective search is evolutionary stable. Also, the evolution of distributed resource sensing could be shown (Hein et al., 2015) with a stable state at a first order phase transition. The transition region enabled the agents to switch between behaviors depending on the resource abundance. An allee-effect for strong alignment depending on the depletion-rate of a resource was discovered (Wood and Ackland, 2007).

None of the above studies has explored the criticality hypothesis in the context of predators. Since predation is considered the main driver for the emergence of group living animals, it is the candidate for a self-organization mechanism to criticality. Those studies which simulated evolution under predation have not analyzed this possibility, i.e. the distance of the final evolutionary stable state to the order transition was not quantified. I will close this gap in the first part of my thesis and start from an already cohesive collective (as shown in Olson et al., 2013, 2016; van der Post et al., 2015; Wood and Ackland, 2007). My main research questions are:

- Are there other effects than the largest responsiveness of the collective at criticality which might be detrimental?
- Is the order disorder transition an evolutionary stable state?

First, I provide an intuitive picture for the criticality hypothesis, i.e. why the susceptibility is expected to be maximal at the transition. Second, I introduce a predator model and a rudimentary prey model with volume exclusion and a continuous order-disorder phase transition. Third, I derive and discuss the susceptibility and the information transfer in the model system. Third, the group structure and the measures typically used in statistical physics to quantify response are computed from numerical simulations. With biological relevant measures the response to the predator is analyzed in detail. And finally, the simulation of individual evolution is analyzed followed by a discussion of the first part of my thesis.

The majority of the outlined work is published as a preprint (Klamser and Romanczuk, 2020).

2 An intuition for the criticality hypothesis

I will try to provide an intuitive understanding for the criticality hypothesis that does not require any prior knowledge of statistical physics and phase transitions. That means, we have to define the response to a signal in general terms and why it is expected to be largest at a phase transitions. The latter is the motivation for the criticality hypothesis: biological system should operate where they respond best to external signals.

Let us first define the response χ of a system by how a macroscopic observable $\langle m \rangle$ changes if a small signal or field h is applied. Now assume the system consists of cooperating units that tend to align with each other, for example magnets or swarming fish. If the noise η that acts on each unit is weak, the cooperating units succeed and reach an ordered state. This can be quantified by an order parameter (e.g. magnetization as in Fig. 2.1). Thus, the macroscopic observable $\langle m \rangle$ is exactly this order parameter. The phase of the system changes from ordered to disordered at a critical noise η_c. An analog interpretation of noise is temperature and an analog to the critical noise η_c is for example the Curie temperature T_c for specific materials. These materials are permanently magnetized below but not above T_c. I will stick to the example of magnets in the following. If we introduce an external field (e.g. an external magnet) the ordered state is not strongly affected because the magnetization is already almost maximal. That means we expect in the ordered state a weak susceptibility (no difference between blue and red arrows for $\eta \ll \eta_c$ in Fig. 2.1). If noise dominates ($\eta \gg \eta_c$) the system is disordered and a weak external field can not induce order. Thus, also in the disordered state we expect a weak susceptibility (no difference between blue and red arrows for $\eta \gg \eta_c$ in Fig. 2.1). At the transition from order to disorder, the noise ($\eta = \eta_c$) is just enough to prevent order, i.e. the system is at the "edge of chaos". A small external field is sufficient for the units tendency to align with each other to dominate the noise. Finally, this leads to the expected peak in susceptibility at the phase transition (difference between blue and red arrows at η_c in Fig. 2.1).

Please note, that the explanation given above is extremely simplified. It does not even take into account the most popular feature at phase transitions, the divergence in correlation length. There are also other features at criticality that depend on the specific system, for example a diverging signal to noise ratio (for a great review on criticality in living systems see Muñoz, 2018). However, I favored the simplicity of the picture above to make the idea quickly accessible, but it should be taken with caution.

Let us move from the realm of physics and magnetization to biology and the polarization of a shoal. In this context, the external field is the signal to which a biological system like a shoal of fish needs to respond to, i.e. it could represent a predator. The criticality hypothesis, which I intend to test in the following chapters, assumes that a collective of interacting agents self-tunes to the phase transition because there it responds strongest to the predator. Thus, we developed a simplified but intuitive understanding for the reasoning behind the hypothesis.

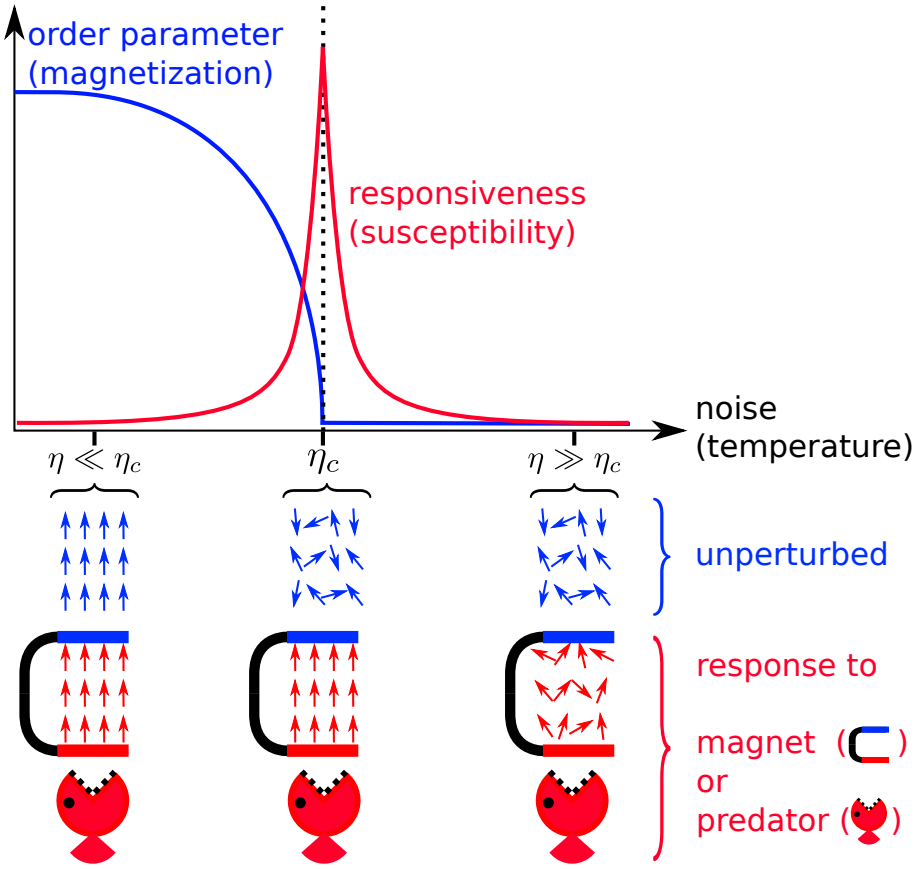

Figure 2.1: Susceptibility at a phase transition. The blue curve represents an order parameter, e.g. magnetization, which decreases with increasing noise η (or temperature). The order parameter vanishes if the noise is larger than the critical noise $\eta > \eta_c$. Therefore, η_c marks a phase transition. The red curve represents the susceptibility χ, which peaks at the phase transition. The susceptibility is the response of the system, e.g. the change in order parameter, to an external field or signal. Thus, an intuitive understanding of why it peaks at the transition is that the ordered state ($\eta \ll \eta_c$) can not much increase in order due to an external field. In the disordered state ($\eta \gg \eta_c$) the noise is so strong that a small external field is not sufficient to change the system state. In contrast, at the transition ($\eta = \eta_c$) the noise is already in balance with the internal tendency to align and a small field can cause the whole system to be polarized. In this thesis the field is assumed to represent the endeavor to avoid a predator.

3 Modeling predator prey interaction

This chapter presents the model of the prey collective and of the single predator. All model parameters are listed in Table 3.1. At the end, I compare it with other models from a conceptual point of view and discuss the generality of the possible outcomes.

3.1 Prey collective

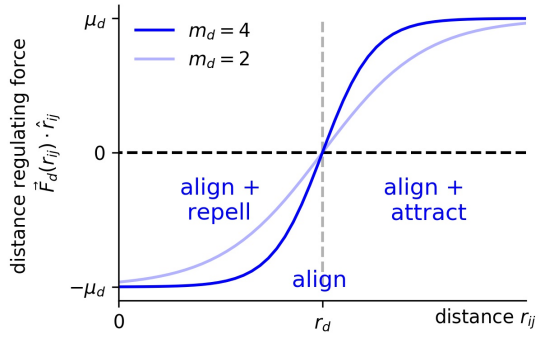

Figure 3.1: Distance regulating force $\vec{F}_d(r_{ij})$ between agents i and j projected on the separation direction $\hat{r}_{ji} = \frac{\vec{r}_j - \vec{r}_i}{|\vec{r}_j - \vec{r}_i|}$. The force equals zero at the favored distance $r_d = 1$ and is displayed for a distance slope $m_d = -2$ (used in the simulations) and $m_d = -4$.

The prey agents are modeled as self-propelled particles with constant velocity $v = v_0$ and angular noise (Romanczuk and Schimansky-Geier, 2011) by solving the following stochastic differential equations:

$$\frac{d\vec{r}_i(t)}{dt} = \vec{v}_i(t) = v_0(\cos(\varphi_i), \sin(\varphi_i)) \quad (3.1a)$$

$$\frac{d\varphi_i(t)}{dt} = \frac{1}{v_0}\left(F_{i,\varphi}(t) + \sqrt{2D}\xi(t)\right) \quad (3.1b)$$

with $F_{i,\varphi}(t) = \vec{F}_i(t) \cdot \vec{e}_\varphi$ as the force acting on agent i in its angular direction $\vec{e}_\varphi$, D as the angular diffusion coefficient and $\xi(t)$ as Gaussian white noise. For simplicity, I omit in the following the explicit time dependence of position, velocity and forces. Agents react to their environment by aligning to their conspecifics, keeping a certain distance to them and by fleeing from a predator. The alignment force between a focal agent i and all its neighbors $j \in \mathbb{N}_i$ tends to minimize the velocity difference $\vec{v}_{ji} = \vec{v}_j - \vec{v}_i$:

$$\vec{F}_{i,a} = \frac{1}{|\mathbb{N}_i|} \sum_{j \in \mathbb{N}_i} \mu_{alg} \cdot \vec{v}_{ji} \ . \quad (3.2)$$

Their desired distance r_d to each other is regulated by the distance regulating force

$$\vec{F}_{i,d} = \frac{1}{|\mathbb{N}_i|} \sum_{j \in \mathbb{N}_i} \mu_d \cdot \tanh\left(m_d(r_{ji} - r_d)\right) \cdot \hat{r}_{ji} \quad (3.3)$$

with $\hat{r}_{ji} = \frac{\vec{r}_j - \vec{r}_i}{|\vec{r}_j - \vec{r}_i|}$ as direction from agent i to j, μ_d as strength of the force and m_d as the slope of the change from repulsion (for $r_{ji} < r_d$) to attraction (for $r_{ji} > r_d$). The force is illustrated in Fig.3.1. I have chosen the hyperbolic tangent (tanh) because it has some interesting properties: (i) It has a sigmoidal character, i.e. it saturates for extreme large and small distances and does not grow infinitely which would be unphysical/unbiological. (ii) It is point symmetric and the corresponding potential is therefore symmetric around r_d. (iii) It is continuous and therefore different from the discrete zone models, where the discreteness can cause unbiological behavior as an oscillatory behavior between strong repulsion and attraction.

Finally, if a predator p is a neighbor of agent i, $p \in \mathbb{N}_i$, the agent is repelled according to

$$\vec{F}_{i,f} = -\mu_{flee} \cdot \hat{r}_{pi} \ . \quad (3.4)$$

The total force acting on agent i is thus defined as

$$\vec{F}_i = \vec{F}_{i,d} + \vec{F}_{i,alg} + \vec{F}_{i,flee} \ . \quad (3.5)$$

	parameter	symbol	value
prey	angular diffusion	D	0.5
	alignment strength	μ_{alg}	evolves
	distance strength	μ_d	2
	distance slope	m_d	-2
	(distance favored)	r_d	1
	(speed)	v_0	1
	flee strength	μ_{flee}	4
predator	speed	v_p	2
	pursuit strength	μ_p	2
	attack rate	γ_a	1/3
	catch radius	r_{catch}	3
	initiation distance	r_{pred}	1.5 r_{catch}
simul.	number of agents	N	400
	time step	dt	0.02
	equilibration time	T_{eq}	200
	simulation time	T_{simu}	120
	mutation rate	γ_m	0.8
	mutation strength	σ_m	0.075

Table 3.1: Parameters used in the predator-prey model and for the evolution. The prey parameters "speed" and "distance favored" can be used in the dimensionless time and length units leaving all the other parameters unchanged. Note that the "flee strength" is strictly speaking a predator-prey interaction parameter which reduces the prey-parameter to effectively four.

3.1.1 Parameter setting

The model parameters are listed in Table 3.1. Note that two parameters can in principle be eliminated by rendering the equations dimensionless. If, for instance, the favored distance r_d and the prey speed v_0 are used to define the characteristic length L and time T:

$$L = r_d,\ T = \frac{r_d}{v_0}\ , \tag{3.6}$$

the Eq.3.1 can be reformulated to

$$\frac{d\vec{r'}_i}{dt'} = \vec{v'}_i \tag{3.7a}$$

$$\frac{d\varphi_i}{dt'} = \frac{r_d}{v_0^2}\left(F_{i,\varphi} + \sqrt{2D}\sqrt{\frac{v_0}{r_d}}\xi(t')\right) \tag{3.7b}$$

$$= F'_{i,\varphi} + \sqrt{\frac{2D_r r_d}{v_0}}\xi(t')\ . \tag{3.7c}$$

Here is $D_r = \frac{D}{v_0^2}$ the rotational diffusion coefficient (with the unit $[D] = 1/t$), the primed variables are the dimensionless counterparts

$$t = \frac{r_d}{v_0}t',\ v_i = v_0 v'_i,\ r_i = r_d r'_i \tag{3.8}$$

and note that the Gaussian process is transformed to

$$\xi(t) = \sqrt{\frac{v_0}{r_d}}\xi(t')\ . \tag{3.9}$$

With this choice of characteristic length and time and setting $v_0 = 1$ and $r_d = 1$, the dimensionless parameters keep their values listed in Table 3.1.

Since the flee strength μ_{flee} is a predator-prey interaction parameter, the prey system has effectively only four parameters from which the alignment strength μ_{alg} is evolving. The remaining prey-parameters are the angular-diffusion coefficient D which is set to $D = 0.5$ resulting in a persistence time of $\tau_p = \frac{v_0^2}{D} = 2$, i.e. an solitary agent keeps its rough direction for two body length. The distance strength $\mu_d = 2$ which sets a force-reference and ensures that agents stay cohesive. The distance slope $m_d = -2$ regulates how quick the distance force saturates to its maximal/minimal value away from the favored distance r_d (Fig.3.1).

The simulation parameters are chosen to be realistic for medium shoal-sizes $N = 400$ and as computationally cheap as possible, the latter is meant to facilitate the reproducibility of this thesis. Note that in each generation 76 independent simulations are run with each equilibrating for $T_{eq} = 200$ and running then for $T_{simu} = 120$ time units. The equilibration time is shown to be sufficient for a collective to relax in its stationary state (Sec. A.1). The time-step is set to $dt = 0.02$ which provides sufficient stability and efficient computation (see Sec. A.2).

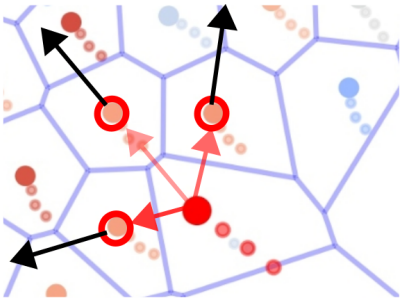

Figure 3.2: Illustration of predator attack. Predator (large red point) pursues its possible targets, i.e. its frontal Voronoi neighbors (smaller points). The red arrows represents the attraction towards its targets with the opacity of the respective strength or weight. The black arrows represent the flee-direction of the targets.

The simulation length T_{simu} was chosen to ensure that the predator can reach the whole collective. A shoal of 400 agents with a preferred distance of $r_d = 1$ should have a characteristic length of about $L \approx 20$. If I assume the predator oscillates between the front and the back in a straight line, which is obviously not the case, it travels to the front with a relative velocity of 1 and to the back with 3. Thus, it could in principle oscillate about 4 times which is sufficient time to encounter the whole swarm.

3.2 Single predator

For simplicity the predator behaves deterministic, i.e. no angular diffusion, which simplifies Eq.3.1b to

$$\frac{d\varphi_p}{dt} = \frac{1}{v_p}\vec{e}_\varphi \cdot \vec{F}_p \tag{3.10}$$

with v_p as the fixed predator speed and $\vec{F}_p$ as the pursuit force. In this study, the predator is faster than the prey $v_p > v_0$. I assume that the predator can only attack one prey at a time, considers its frontal Voronoi-neighbors $\mathbb{N}_p$ as targets and selects equally likely among them:

$$p_{select,i} = \begin{cases} \frac{1}{|\mathbb{N}_p|} & \text{if } i \in \mathbb{N}_p \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

The limitation that targets are only frontal Voronoi-neighbors $\mathbb{N}_p$ represents the belief that predators select targets who are not occluded and whose position and speed are accessed by binocular vision. If the predator launches an attack, with an attack rate γ_a which also represents handling time, it's success probability is linear distance dependent and zero at a distance larger than r_{catch}:

$$p_{success,i} = \min\left(\frac{r_{catch} - r_{ip}}{r_{catch}},\ 0\right) . \tag{3.12}$$

Summarizing, the probability that a predator successfully catches a targeted agent within a small time window $[t, t + \delta t]$ is

$$p_{catch,i}(t, \delta t) = p_{success,i}(t) \cdot p_{select,i}(t) \cdot \gamma_a \delta t \tag{3.13}$$

The predator aims at the weighted center of mass of the perceived prey (Fig. 3.2). Each prey position is weighted by its probability of a successful catch $p_{catch,i}(t, \delta t)$. Since $p_{catch,i}$ is non-zero

only for the predator's frontal Voronoi-neighbors, the predator effectively tries to maximize its catch based on local information. The pursuit force is thus

$$F_p = \mu_p \cdot \left(\sum_i p_{catch,i} \hat{r}_{ip} \right) \tag{3.14}$$

with $\hat{r}_{ip}$ as the direction vector pointing from the predator p to prey agent i.

3.2.1 Initialization or predator appearance

Since the prey can sense the predator based on a topological neighborhood criterion, they would flee irrespective of the separating distance. This model assumption is unrealistic because the ability to perceive objects has a fundamental limit (cycles per degree; Pita et al., 2015) and even if fish can perceive the stimulus it is known from social interactions in between conspecifics that the interaction strength decreases with distance (Rosenthal et al., 2015). However, if the predator is in general close this assumption is justified. Therefore, it is important that the predator is initialized not too far away from the school but within a realistic reaction distance from the closest individuals. In the simulations, the predator is initiated at a distance of $d_{p,0} = 1.5 r_{catch}$ away from the alpha shape of the shoal. The radius used to compute the alpha shape is $r_\alpha = 6 r_d$. With this choice of r_α, the structure of the school is taken into account, while avoiding a rough surface that would effectively lower the initiation distance of the predator.

3.2.2 Parameters

The predator speed must be larger than the prey-speed and is set to $v_p = 2$. Its pursuit strength μ_p describes together with the speed the predator's turning ability and is set to $\mu_p = 2$ and therefore equals the preys distance force strength. With a capture rate $\gamma_c = 1/3$ and a simulation time of $T = 120$, around forty preys are caught per round, i.e. ten percent of the shoal. The catch radius is set to $r_{catch} = 3$ (i.e. three body length).

3.3 Model comparison

In this section, I compare the introduced prey model to other prominent flocking models from the literature.

The most prominent and simple model is the Vicsek-model (Vicsek et al., 1995). It is a rule based model where agents align their heading direction φ with the agents that are closer than a threshold distance

$$\varphi_i(t+1) = \langle \varphi(t) \rangle_{N_r} + \Delta\varphi \ . \tag{3.15}$$

Here $\langle \dots \rangle_{N_r}$ is the average over all agents that are closer than r to the focal agent i. $\Delta\varphi$ is a random variable drawn uniformly from $[-\eta/2, \eta/2]$. Since there is no attraction/repulsion force, studies on this system normally apply periodic boundary conditions, otherwise the "swarm" would quickly disperse. Another difference to the prey model (Sec.3.1) is that the interactions between agents are based on a purely metric criterion and therefore lead to a strong feedback between density and order: in regions of higher density the number of interaction partner increases which averages out directional noise and leads to a stronger polarization. This leads to the formation of bands in the transition region (Chaté et al., 2008). Bands are elongated regions with high density and high order moving perpendicular to their long-axis through regions of low density and order. They are an example of phase coexistence which reveals the discontinuous character of the order-disorder transition in the Vicsek model. Importantly, this phase coexistence at the transition renders the Vicsek model unsuitable as a candidate to test the criticality hypothesis (it lacks properties such as divergent susceptibility, i.e. maximal responsiveness). The lack of a repulsive force, which can be considered nonphysical because it corresponds to point-particles, even sets the

Vicsek model no upper bound for the density (most probably enhancing the feedback between density and order). However, with volume exclusion, by the introduction of a repulsive force, the discontinuous character of the phase transitions remains (Chaté et al., 2008). Note that initially (Vicsek et al., 1995) the order transition in the Vicsek model was assumed to be continuous and only at larger system sizes (Chaté et al., 2008) its discontinuity was revealed. Therefore, the finite size of a system, certainly fulfilled in natural swarms, could lead to quasi second-order transition phenomena.

For a topological version of the Vicsek model the nature of the transition is continuous as shown by Ginelli and Chaté (2010). They used Voronoi interactions which effectively decoupled the local density from the order. With metric interactions agents in high densities have a larger number of neighbors and therefore average out fluctuations, i.e. increasing the order. This is not the case for agents with Voronoi interactions. Since also the prey model (Sec.3.1) is based on a Voronoi tessellation, the order-disorder transition is of the same continuous nature.

Another prominent model is the three-zone model (Couzin et al., 2002) which has in addition to the alignment and repulsion zone, a third attraction zone. The model was adapted in Couzin et al. (2005) such that the attraction and alignment zone overlap with each other (applied also in: Couzin et al., 2011; Guttal and Couzin, 2010). The interplay of attraction and repulsion results in cohesive swarms, which is the typical form of collective behavior in animal groups (Krause and Ruxton, 2002). With cohesion comes a difference in relative position with respect to the center of mass, the movement direction and/or the edge of the collective. If heterogeneity is introduced, for example via parameters as speed, turning rate, repulsion range, a self-sorting of individuals with respect to the center of mass or the front is expected (Couzin et al., 2002; Hemelrijk and Kunz, 2005; Jamie Wood, 2010). Heterogeneity plays an important role in collective behavior (Jolles et al., 2017) and it is fundamental to evolutionary studies. The prey model (Sec.3.1) has also two distinct zones (attraction and repulsion) and an overlapping alignment zone. Therefore, it is as topological as a model with cohesion and volume exclusion can be: the attraction and alignment are not distant dependent (for $r \gg r_d$) and the only length scale is the preferred distance r_d at which the distance regulating force switches from repulsion to attraction. This reduces the number of parameters but an infinite sensory range is unrealistic. However, since the equilibrated state is a cohesive state, the distance between Voronoi neighbors is limited in the bulk to short ranges. Only at the edge of the swarm are interactions over a greater distance possible. Still, their range is limited by the diameter of the school.

4 Susceptibility and information transfer

A quantity of focal interest in studies concerned with phase transitions (Barnett et al., 2017; Calovi et al., 2018; Kyriakopoulos et al., 2019) is the susceptibility, which describes the responsiveness of the system to external signals. In this chapter, I discuss the approximation of susceptibility and information transfer in the proposed model system. Since I highlight links to statistical physics, the following sections might be difficult to follow for a reader unfamiliar to this topic. However, I tried to frame everything in terms of probabilities, which allows a reasonable understanding also without corresponding background. The susceptibility and its estimation via thermodynamic relations is described for the Ising model, which is a model to understand phase transitions in a simplified magnetic system. I analogously derive the susceptibility for the prey model under the assumption that the system is in thermodynamic equilibrium. I examine if the analysis of steady states instead of equilibrium states via the maximum entropy approach is possible. The discussion of susceptibility is closed by reviewing the basic assumptions that are likely violated when used to estimate the response to a predator. Second, I discuss the possibility to measure the information transfer via local velocity fluctuations and examine as well the limitations of this approach.

4.1 Susceptibility

4.1.1 Susceptibility in the Ising model

The susceptibility is in general defined by how strong a macroscopic observable $\langle m \rangle$ changes if an external field h is changed

$$\chi = \frac{\partial \langle m \rangle}{\partial h} \ . \tag{4.1}$$

In the Ising-model, the susceptibility defined in Eq. 4.1 describes the change of the magnetization per spin

$$m = \frac{M}{N} = \frac{1}{N} \sum_{i=1}^{N} s_i \ , \tag{4.2}$$

given the change of an external field h The s_i is the spin at side i which can be either up or down, i.e. $s_i \in [-1, 1]$. Interestingly, this directly links the reaction of the system to fluctuations in the order parameter. This can be shown with tools from statistical physics of systems in thermodynamic equilibrium. In statistical physics the probability to observe the system in the state $\vec{s} = [s_0, s_1, \ldots, s_N]$ is

$$P(\vec{s}) = \frac{\exp[-\beta H(\vec{s})]}{Z} \ . \tag{4.3}$$

$H(\vec{s})$ describes the energy of the system at state $\vec{s}$ and β is the inverse of the thermal energy $\beta = 1/(k_b T)$ with k_b as the Boltzmann constant and T as the temperature of the surrounding heat bath. Thus, the state $\vec{s}$ is more likely the smaller its corresponding energy. The partition function

$$Z = \sum_{\{\vec{s}\}} \exp[-\beta H(\vec{s})] \tag{4.4}$$

normalizes the probability with $\sum_{\{\vec{s}\}}$ as a sum over all possible system state. If spins tend to align with the external field, the energy is partly defined as $H(s_i) = ... - h \sum_i s_i$. Now, the mean

magnetization per spin can be computed to

$$\langle m \rangle = \sum_{\{\vec{s}\}} m(\vec{s}) P(\vec{s}) = 1/N \frac{1}{\beta} \frac{\partial \ln Z}{\partial h} \ . \tag{4.5}$$

This allows us to derive the susceptibility χ defined in Eq. 4.1 to

$$\chi = \frac{1}{\beta} \frac{\partial^2 \ln Z}{\partial h^2} = \frac{\beta}{N} [\langle M^2 \rangle - \langle M \rangle^2] = \beta N [\langle m^2 \rangle - \langle m \rangle^2] \tag{4.6}$$

The above relation is also referred to as fluctuation dissipation theorem, because it relates fluctuation in the magnetization with the thermal energy (remember β is the inverse of the thermal energy and therefore represents the dissipation). It connects the response of the system to an infinitesimally small change of the external field h with fluctuations in the order parameter. The linear nature of this response to small changes can also be assessed by a Taylor-expansion to linear order of the canonical distribution around $h = 0$ (see for example Eq.1.21 in Marconi et al., 2008). The response can be reformulated to highlight the link to the connected spin correlation function or spin pair correlation function

$$\chi = N\beta [\langle m^2 \rangle - \langle m \rangle^2] = \frac{\beta}{N} \left[\left\langle \sum_{ij} s_i s_j \right\rangle - \left\langle \sum_i s_i \right\rangle \cdot \left\langle \sum_j s_j \right\rangle \right] \tag{4.7a}$$

$$= \frac{\beta}{N} \sum_{ij} [\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle]. \tag{4.7b}$$

In the following, I establish an analog description for the model system (presented in Sect. 3.1) with fixed speed.

4.1.2 Susceptibility of the prey collective in equilibrium

For simplicity I assume, as in the section before, that the prey agents (Sect. 3.1) react to a global homogeneous field $\vec{h}$. From Eq. 3.1 the change in heading of individual i in response to $\vec{h}$ is

$$\frac{d\varphi_i}{dt} = \frac{\vec{h}\hat{e}_{\varphi,i}}{v_0} = F_{i,s} \text{ with } \hat{e}_{\varphi,i} = [-\sin\varphi_i, \cos\varphi_i] \ . \tag{4.8}$$

From this force $F_{i,s}$ the analog to energy $H_{s,i}$ for individual i can be computed via integration to

$$H_{s,i} = -\frac{\vec{h}\hat{u}_i}{v_0} \text{ with } \hat{u}_i = [\cos\varphi_i, \sin\varphi_i] \ . \tag{4.9}$$

The total energy is composed of the sum of isolated components $H_{s,i}$ and of the part that is influenced by the interactions in between the prey H_m:

$$H = H_m(\vec{\varphi}) + \sum_i H_{s,i}(\varphi_i, \vec{h}) = H_m(\vec{\varphi}) + -\frac{\vec{h}}{v_0} \cdot \sum_i \hat{u}_i \tag{4.10}$$

with $\vec{\varphi} = [\varphi_0, \varphi_1, \ldots, \varphi_N]$. Only $H_{s,i}$ depends on the external field $\vec{h}$. Knowing the energy of the systems allows (analog to Eq. 4.3) to define a probability to observe the state $\vec{\varphi}$ which is

$$P(\vec{\varphi}) = c_H \frac{\exp[\beta \vec{h} \sum \hat{u}_i]}{Z} = c_H \frac{\exp[\beta N \vec{h} \vec{\phi}]}{Z} \tag{4.11}$$

with $c_H = e^{-\beta H_m}$. However, note that Eq. 4.3 assumes that there is a heat bath represented by $\beta = 1/(k_b T)$. Since the strength of the angular noise D (see Eq. 3.1) can prevent polarization in

the prey collective, it plays a similar role as the temperature in the Ising model. Therefore, I use $\beta = 1/(Dv_0)$ to compute the expectation value of the polarization vector $\vec{\Phi} = \frac{1}{N}\sum_i^N \hat{u}_i$ (analog to the computation of the mean magnetization in the Ising model).

$$\left\langle \vec{\Phi} \right\rangle = \sum_{\{\vec{\varphi}\}} \vec{\Phi} P(\vec{\varphi}) = \frac{1}{N\beta} \vec{\nabla}_{\vec{h}} \ln Z \tag{4.12a}$$

$$= \frac{1}{N\beta} \begin{pmatrix} \frac{\partial}{\partial h_x} \\ \frac{\partial}{\partial h_y} \end{pmatrix} \ln \left(\sum_{\{r,\varphi\}} c_H e^{\beta \vec{h} \cdot \vec{M}} \right) \ , \tag{4.12b}$$

with $\vec{M} = N\vec{\Phi}$. Finally, I compute the susceptibility as the sum of changes of the polarization vector $\left\langle \vec{\Phi} \right\rangle$ components with respect to the external field $\vec{h}$. It can be written more compact with the $\vec{h}$-Laplace operator $\Delta_{\vec{h}} = \frac{\partial^2}{\partial h_x^2} + \frac{\partial^2}{\partial h_y^2}$ to

$$\chi = \vec{\nabla}_{\vec{h}} \left\langle \vec{\Phi} \right\rangle = \frac{1}{N\beta} \Delta_{\vec{h}} \ln(Z) \tag{4.13a}$$

$$= \frac{\beta}{N} \left[\left\langle M_x^2 + M_y^2 \right\rangle - \left(\langle M_x \rangle^2 + \langle M_y \rangle^2 \right) \right] \tag{4.13b}$$

$$= \frac{\beta}{N} \left[\left\langle \vec{M} \cdot \vec{M} \right\rangle - \left\langle \vec{M} \right\rangle \cdot \left\langle \vec{M} \right\rangle \right] \tag{4.13c}$$

$$= \beta N \left[\left\langle \vec{\Phi} \cdot \vec{\Phi} \right\rangle - \left\langle \vec{\Phi} \right\rangle \cdot \left\langle \vec{\Phi} \right\rangle \right] = \beta N \left[\left\langle \Phi^2 \right\rangle - \langle \Phi \rangle^2 \right] \ . \tag{4.13d}$$

This is analogous to Eq. 4.7 and establishes a link to the pair-correlation between individual heading direction. Analogously to Eq. 4.7, we may also write:

$$\chi = N\beta \left[\left\langle \vec{\Phi} \cdot \vec{\Phi} \right\rangle - \left\langle \vec{\Phi} \right\rangle \cdot \left\langle \vec{\Phi} \right\rangle \right] \tag{4.14a}$$

$$= \frac{\beta}{N} \left[\left\langle \sum_i \hat{u}_i \cdot \sum_j \hat{u}_j \right\rangle - N^2 \left\langle \vec{\Phi} \right\rangle \cdot \left\langle \vec{\Phi} \right\rangle \right] \tag{4.14b}$$

$$= \frac{\beta}{N} \left[\left\langle \sum_{ij} \hat{u}_i \cdot \hat{u}_j \right\rangle - \sum_{ij} \left\langle \vec{\Phi} \right\rangle \cdot \left\langle \vec{\Phi} \right\rangle \right] \tag{4.14c}$$

$$= \frac{\beta}{N} \sum_{ij} \left[\langle \hat{u}_i \cdot \hat{u}_j \rangle - \left\langle \vec{\Phi} \right\rangle \cdot \left\langle \vec{\Phi} \right\rangle \right] \tag{4.14d}$$

$$= \frac{\beta}{N} \sum_{ij} \left\langle \left(\hat{u}_i - \left\langle \vec{\Phi} \right\rangle \right) \cdot \left(\hat{u}_j - \left\langle \vec{\Phi} \right\rangle \right) \right\rangle \ . \tag{4.14e}$$

Note that the above derivation until Eq. 4.13 assumes a thermodynamic equilibrium and is for the out-of-equilibrium prey model strictly speaking not valid. However, from Eq. 4.13 to Eq. 4.14 nothing like that is assumed, it is just a reformulation and therefore valid. It means, that we can interpret χ always by referring to the sum over all pair correlation in velocity fluctuations.

With the above considerations we can revisit the statement: At a continuous phase transition the susceptibility is maximal at zero external field. If a small field h is applied, the system relaxes to an equilibrium state. The difference of this state compared to the equilibrium state before the field h was present is maximal at the phase transition. In the context of prey agents the change in polarization is of interest.

4.1.3 Maximum entropy approach: susceptibility of steady states

In the above derivation of the susceptibility, I used concepts of equilibrium statistical physics. However, the system is out of equilibrium (see Sect.4.1.4) and therefore changes in its steady

states should be analyzed instead. I will sketch how the maximum entropy approach could be applied to the system to define a probability distribution of system states. The only condition that needs to be fulfilled is the stationarity of the system; it should be in a stable state. The basic idea, as described e.g. by Bialek et al. (2014), is that the observables of the system are described by the most simple probability distribution which is quantified by the Shannon entropy of the distribution

$$S(P) = -\sum_{\{\vec{\varphi}\}} P(\vec{\varphi}) \ln P(\vec{\varphi}) \ . \tag{4.15}$$

It means the distribution with the largest Shannon entropy is the simplest one. Since the distribution also needs to result in the experimental observables

$$\langle O_\mu \rangle = \sum_{\{\vec{\varphi}\}} P(\vec{\varphi}) O_\mu = \langle O_\mu \rangle_{exp} \tag{4.16}$$

these has to be included in the generalized entropy with Lagrange multipliers h_x and h_y for the observables Φ_x and Φ_y respectively. Note, experimental observables would correspond to numerical simulations of the model. The additional constrain that the probability distribution sums up to 1, is taken into account by the Lagrange multiplier λ_0. This results in

$$S_g(P, h_x, h_y) = S(P) - \sum_{i \in \{x,y\}} h_i \left[\langle \Phi_i(\vec{\varphi}) \rangle_P - \langle \Phi_i(\vec{\varphi}) \rangle_{exp} \right] - \lambda_0 \left[\langle 1 \rangle_P - 1 \right] \tag{4.17}$$

$$= S(P) - \vec{h} \cdot \left[\left\langle \vec{\Phi} \right\rangle_P - \left\langle \vec{\Phi} \right\rangle_{exp} \right] - \lambda_0 \left[\langle 1 \rangle_P - 1 \right] \ . \tag{4.18}$$

Note that it is possible to take the x and y component of the polarization, despite their obvious correlation, separately into account. The observables are always computed from the detailed state of the system $\vec{\varphi}$ and therefore always obey the inequality $\sqrt{\Phi_x^2 + \Phi_y^2} \leq 1$. The two constraints are also linear independent because only $\langle \Phi_x \rangle_{exp} = 1$ defines $\langle \Phi_y \rangle_{exp}$ to zero.

To fulfill the simplicity criterion the generalized entropy Eq. 4.17 needs to be maximized with respect to $P(\vec{\varphi})$:

$$\frac{\partial S_g}{\partial P(\vec{\varphi})} = 0 \tag{4.19a}$$

$$= -\sum_{\vec{\varphi}} \left[\ln P(\vec{\varphi}) + 1 + \vec{h} \cdot \vec{\Phi} + \lambda_0 \right] \ . \tag{4.19b}$$

Reformulating with respect to P leads to

$$P(\vec{\varphi}) = \exp\left[-(1 + \lambda_0)\right] \exp\left[-\vec{h} \cdot \vec{\Phi}\right] \ . \tag{4.20}$$

The optimization, as done above, with respect to λ_0 normalizes the distribution

$$P(\vec{\varphi}) = \frac{\exp\left[-\vec{h} \cdot \vec{\Phi}\right]}{Z(\{\vec{\varphi}\})} \tag{4.21}$$

with $Z(\{\vec{\varphi}\}) = \sum_{\{\vec{\varphi}\}} \exp\left[-\vec{h} \cdot \vec{\Phi}\right]$.

With this probability density arbitrary expectation value $\langle f(\vec{\varphi}) \rangle$ of a function which depends only on $\vec{\varphi}$ can be computed, as for example the average polarization $\langle \Phi \rangle$. Of course, $\langle \phi \rangle$ will correspond to the experimental observable $\langle \phi \rangle_{exp}$ since that is how the distribution was derived. In addition, expectation values that have not been used to derive the distribution are computable. If we condition the probability to also reflect for example the nearest neighbor distance, also expectation values based on relative positions of preys are computable. However, the maximum entropy

approach is not explicitly linked to any model parameter; it is purely based on the distribution generated from experimental data. The Lagrangian multiplier might have a physical meaning but it is not plausible to assume that for example $\vec{h}$ represent an external field. Thus, it is not possible to compute the susceptibility via Eq. 4.1 and therefore no fluctuation dissipation theorem is available for stationary non-equilibrium systems via this approach. There are other approaches to non-equilibrium systems than the maximum entropy approach that might offer an opportunity (for a recent review see Sarracino and Vulpiani, 2019). Especially the assumption of a small perturbation that allows the application of linear response theory is promising to derive a relation for the susceptibility (Marconi et al., 2008).

If the probability distribution of the microscopic states can be derived from the model, the susceptibility could be computed for the system. But then the maximum entropy approach would not be necessary anymore since the distribution is already known.

4.1.4 Difference between susceptibility and predator response

I assumed in Sect.4.1.2 that (i) the system is in thermodynamic equilibrium (ii) the changes of the external field are small and it is (iii) global and (iv) homogeneous. These four are in general violated for the reaction of a collective to a predator.

- **Equilibrium state**: The system is an active system and therefore per definition a non-equilibrium system. The agents dissipate constantly energy (no conservation of momentum) but, due to an unspecified energy source, keep their preferred speed, i.e. the system is out of thermal equilibrium.
- **Small changes** of an external field: In the context of a predator attack, the perturbing force is the flee-force of the agent. This flee-force is far from small and normally dominates all other forces. Therefore, to compute the susceptibility by the linear approximation might not be justified.
- **Global field**: The global homogeneous field simplified the former analytical derivations of the susceptibility. However, the flee-force is neither global nor homogeneous. The flee-force acts only on agents that directly sense the predator. If I assume visual interactions with occlusion by conspecifics, but also with metric-, Voronoi-interaction and other local interaction types, the predator is per definition a local perturbation.
- **Homogeneous field**: The flee-force is in the simplest case a repulsion force and therefore inhomogeneous. However, close individuals have similar relative position with respect to the predator and therefore also a similar flee-force. Thus, locally the force can be approximated to be homogeneous.

The violation of the first assumption means that I can not ensure that the fluctuations in the order parameter represent the response of the system to an external field. However, as shown in Eq. 4.14 these fluctuations are analog with the sum over all pair correlations of velocity fluctuations (possible measure for information transfer as discussed in Sec. 4.2). Furthermore, even if I assume that the susceptibility would represent the change of one non-equilibrium stationary state to another one due to an external field, it might be useless at the phase transition. Phase transitions are up to a certain degree analogous to bifurcations in dynamical systems, i.e. both mark the sudden emergence or extinction of steady states. Thus, as it is typical for bifurcations, also at phase transitions critical slowing down occurs. This means that the dynamic of the system slows down and the relaxation to the steady state takes longer the closer the system is to the phase transition. The attack of a predator is fast and the predator does not wait for the collective to reach a steady state to continue. Critical slowing down might be beneficial in other context: the collective stays for a longer time in a perturbed state and therefore has a longer collective memory. However, in the predator-avoidance context it is an additional reason, with the other mentioned unmet assumptions, why the susceptibility should be considered with caution and why its link to optimal predator response is unclear.

4.2 Local information transfer

A possible way to quantify information transfer is by measuring the correlation length in the system as discussed in Cavagna et al. (2010). They argue that scale-free correlation is necessary to transmit undamped information, because otherwise only a perturbation wave goes through the collective that changes the state of the agents on its way but the information is modified during the propagation as well. Thus, if the correlation is scale-free, all agents adapt the same information. This view of information transfer is of a global nature. Also the susceptibility is a global measure and, as demonstrated in Eq.4.14, it can be reformulated to the average correlation of velocity fluctuations between all pairs. This means that the susceptibility is not equivalent to the correlation length but it is proportionate to it.

Therefore, by quantifying the susceptibility I already account for the global nature of the information transfer. Instead of further elaborating in this direction, by also analyzing the correlation length to quantify how local or global the information propagation is, I focus on the strength of local information transfer. This is reasonable in the context of predator-prey interactions because the predator can be located inside the swarm. In consequence does social information, encoded in the flee-direction of individuals, depend on the relative position and can be in conflict with information from different fleeing agents.

4.2.1 Correlation of velocity fluctuations

The most local interactions are those in between neighboring pairs ($\mathbb{N}_i$ as the set of nearest neighbors of agent i) and the information transfer is quantified via the average of the velocity fluctuation correlations

$$C(\delta\vec{v}_i, \delta\vec{v}_j) = \sum_{i,j\in\mathbb{N}_i} \delta\vec{v}_i \cdot \delta\vec{v}_j \ . \tag{4.22}$$

The velocity fluctuation is $\delta\vec{v}_i = \vec{v}_i - \langle\vec{v}\rangle$. Note that this measures not only how correlated but also how strong the fluctuations are. For example, the fluctuations in the disordered phase are large because the group velocity is $\langle\vec{v}\rangle \approx \vec{0}$, and small in the ordered phase. At the same time, the correlation is weak in the disordered phase, because agents do not/weakly align, and strong in the ordered phase. To only account for the correlation the normalized correlation should be used

$$C_N(\delta\vec{v}_i, \delta\vec{v}_j) = \frac{\sum_{i,j\in\mathbb{N}_i} \delta\vec{v}_i \cdot \delta\vec{v}_j}{1/N \sum_i (\delta\vec{v}_i)^2} \ . \tag{4.23}$$

Note that based on the normalized correlation of velocity fluctuations a normalized susceptibility can be computed, by summing over all pairs (as done in e.g. Attanasi et al., 2014a,b). However, as pointed out by Cavagna et al. (2018), this normalized susceptibility would differ from the one derived in statistical physics and therefore would be difficult to interpret because it also depends on the velocity and the order of the system. Also note that the normalized correlation of velocity fluctuations is similar but not equal to the Pearson correlation.

4.2.2 Pitfalls and limitations of correlations as information transfer

I want to stress that the correlation of velocity fluctuation is only a proxy of information transfer.

- The correlation compares velocity at the same moment in time. Per definition, this can not be information transfer, because the information needs time to be copied by an interaction partner. However, if the sampling rate is larger than the smallest time step in the real underlying dynamics, the instantaneous correlation can be caused by an information transfer from one agent to the other and therefore approximates it. One major limitation of this approach is that it does not allow the distinction between transmitter and receiver. To enable this distinction the lagged correlation can be analyzed (Nagy et al., 2010).

- The detected correlations can be due to real interactions in between agents or due to spurious correlation, i.e. both agents respond to a common driver but not to each other. This problem can be taken into account if partial correlations are considered, which are the correlations between residuals. Thus, if we want to investigate if the processes X and Y are correlated given the process Z, we need to estimate the regression coefficient of the former two depending on Z and then correlates their residuals.

- The interactions are nonlinear and linear correlation cannot capture the relationship between the two, or only partially. In this case, information theory can be used to define analogs to the lagged correlation as the transfer entropy (Schreiber, 2000).

These are the obvious problems and limitations if information transfer is approximated by non-lagged correlations of velocity fluctuations. Nevertheless, the simplicity of the measure allows for direct interpretation that would be difficult with other methods. Also in the considered predator prey system some of the above named problems are negligible up to a certain degree.

Regarding the second problem of potential spurious correlations: Since the correlation is computed between nearest neighbors and therefore interacting partners, it should not be a spurious correlation.

The third point, the non-linearity of the interaction process, can also be relaxed. As derived in Sect. A.2, the possible direction changes can be approximated by linear processes. Therefore, the correlation is a valid proxy.

Lastly, I would like to refer the growing literature and tools to measure the transfer of information in terms of inferring causation (e.g. Runge et al., 2019a,b). In general, these techniques are meant for systems where we do not know the underlying relations (as in the earth system on the large scale). However, since the dynamics that govern the system are known, it would be a disproportional effort to apply these tools.

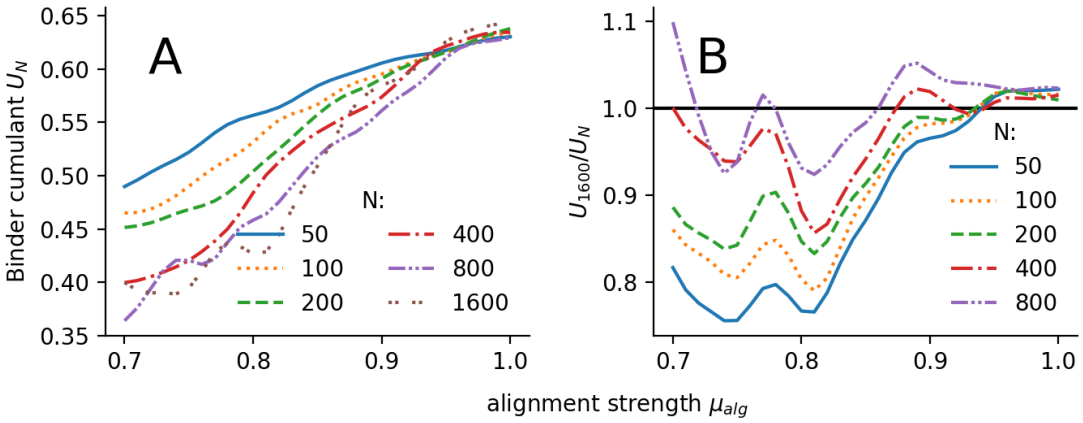

Figure 5.1: Estimation of phase transition via Binder cumulants. Different Binder cumulants U_N are shown for systems varying in sizes as indicated in panel **A**. The phase transition is located where the ratio between Binder cumulants of different sizes becomes one (**B**). The ratio is computed by comparing cumulant of the largest system of $N = 1600$ particles with the smaller systems. The parameters used in the simulations are listed in Tab. 3.1. The simulation lasted for $t = 120$ time units and consists of twenty samples per data point.

5 Group level response

In this chapter, I investigate the collective responsiveness in different ways, i.e. based on fluctuations and on the response to a predator. However, not only the responsiveness but also the spatial shoal structure can influence the group performance. First, I quantify the collective states of the model in terms of structure and polarization and define how the position of the order-disorder transition is estimated. Second, the susceptibility and information transfer are estimated based on intrinsic fluctuation in the collective (both measures have been introduced and theoretically discussed in Sect. 4). Third, the performance of the collective is measured in the presence of a predator and compared to a collective that does not to the predator. Thereby, a biological relevant response variable is used in addition to more *conventional* distance measures. Forth, I test how the violation of the small force assumption (for the computation of the susceptibility via the fluctuation dissipation theorem) changes the location of the optimal response. Additionally, the effect of other obvious parameters that might influence the response are tested (predator speed, steepness of the distance regulating potential). In the final part, the results are discussed.

5.1 Collective states

5.1.1 Order-disorder transition

The phase transition is located where the polarization

$$\Phi = |\langle \vec{u}_i \rangle| \ , \tag{5.1}$$

defined as the absolute value of the average direction vector $\vec{u}_i = \vec{v}_i/|\vec{v}_i|$, becomes different from zero in the thermodynamic limit ($N = \infty$).

To estimate the transition point in a finite system the crossing of the Binder cumulants

$$U_N(\mu_{alg}) = 1 - \frac{\langle \Phi^4 \rangle}{3\langle \Phi^2 \rangle^2} \tag{5.2}$$

of two different system sizes can be used (Binder and Heermann, 2019; Chaté et al., 2008). Therefore, the ratios between Binder cumulants of different system sizes should all cross one at the same critical alignment strength $\mu_{alg,c}$.

However, as shown in Fig. 5.1B, the crossing is not unique and especially for small system sizes it shifts to larger alignment strength. This inconsistency for small system sizes is also reported in the literature (Binder and Heermann, 2019, pp. 50, 107) but could also be due to a low number of samples. However, I do not intend to estimate the critical exponents of the system for which the exact location of the transition is crucial. Thus, a rough estimate of $\mu_{alg,c} = 0.85$ is sufficient which is based on the ratio of the here used largest system size ($N = 1600$, probably the most reliable) and the system size used in the remaining simulations $N = 400$.

The critical value of $\mu_{alg,c} = 0.85$ corresponds to a Binder cumulant of $U_{400} \approx 0.55$ (see Fig. 5.1A) which is used in the rest of the thesis to estimate the transition point in simulation that vary from the standard parameter set (i.e. different to the ones listed in Tab. 3.1). This results in a fair estimation of the transition for different angular diffusion strength (Fig. 5.2).

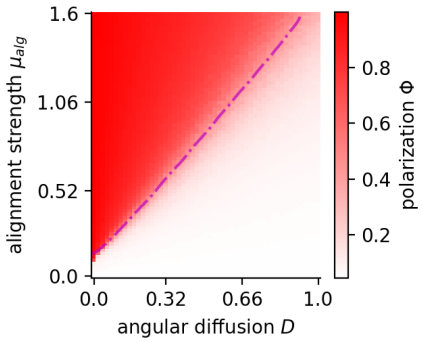

Figure 5.2: Polarization for varying strength of alignment μ_{alg} and angular diffusion D. Dash dotted magenta line marks the order-disorder transition (at Binder cumulant $U_{400} \approx 0.55$).

5.1.2 Structure by density

I will analyze the density via different estimation methods. A local density estimate is based on a K-nearest neighbor approximation ρ and a global estimate on the inverse of the area of the convex hull ρ_{CH}. It is expected that the density estimate of the convex hull ρ_{CH} is lower than the average local estimate ρ if the form of the shoal deviates from a convex hull and if there are vacuoles (void areas) in the shoal, i.e. if the agents are distributed heterogeneously. The vacuoles do not decrease the average local density estimate $\langle \rho \rangle$ because it averages over the densities around fish. Since more fish are present in high density regions and almost none in vacuoles, local density $\langle \rho \rangle$ is a biased approximation.

The local density of agent i is computed with its distance to the kth nearest neighbor $r_{i,kN}$ to

$$\rho_i = k/A(r_{i,kN}, r_{i,e}) \ . \tag{5.3}$$

The term $A(r_{i,kN}, r_{i,e})$ represents the corrected area. If the agent's distance to the edge of the collective $r_{i,e}$ is larger as $r_{i,kN}$, no correction is needed and the area is the area of a circle with radius $r_{i,kN}$. If the distance to the edge is smaller than $r_{i,kN}$, the circle-area is corrected by subtracting the area of the circle segment with a sagitta (height) of $h = r_{i,kN} - r_{i,e}$. Therefore, the area computes to

$$A(r_{i,kN}, r_{i,e}) = \begin{cases} \pi r_{i,kN}^2 & \text{if } r_{i,kN} < r_{i,e} \\ \pi r_{i,kN}^2 - r_{i,kN}\left(r_{i,kN} \arccos \frac{r_{i,e}}{r_{i,kN}} - r_{i,e}\sqrt{1 - \frac{r_{i,e}^2}{r_{i,kN}^2}}\right) & \text{otherwise.} \end{cases} \tag{5.4}$$

This correction is good if the edge of the collective has a small local curvature compared to the curvature of the circle with radius $r_{i,kN}$. This should be fulfilled because a collective of $N = 400$ individuals with a preferred distance of $r_d = 1$ and a spherical form has a radius of $R \approx 11$ while

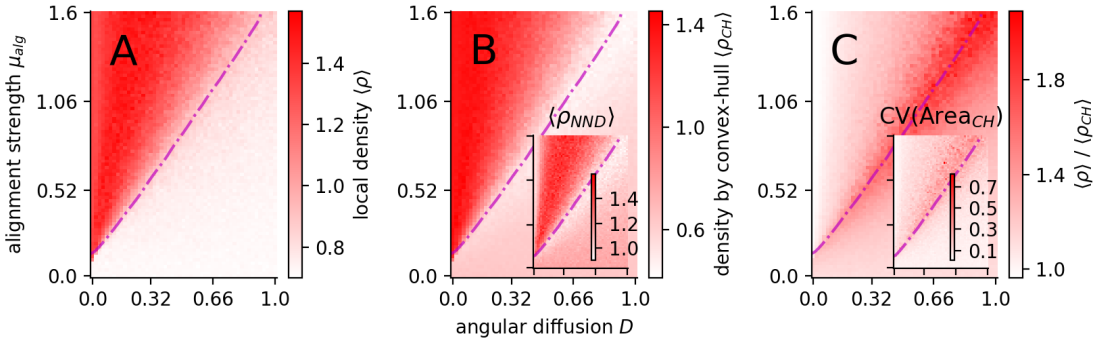

Figure 5.3: Local and global density estimates. **A**: average local density $\langle\rho\rangle$ estimated by the ten nearest neighbors of an agent. **B**: density estimated by number of agents divided by the area of the convex hull $\langle\rho_{CH}\rangle$ and the most local density estimate based on the nearest neighbor distance $\langle\rho_{NND}\rangle$ (inset). **C**: ratio between the different density estimates $\langle\rho\rangle/\langle\rho_{CH}\rangle$ and the coefficient of variation of the area of the convex hull $CV(\text{Area}_{CH})$ (inset). All panels: the dash dotted magenta line marks the order-disorder transition.

the distance to the kth nearest neighbor with $k = 10$ and a Voronoi-interaction network is between 1 and 2.

The global density estimate via the area of the convex hull is computed by

$$\rho_{CH} = \frac{N}{A_{CH}} \,. \tag{5.5}$$

The average of the local density $\langle\rho\rangle$ increases with the order parameter (compare Fig. 5.2 and Fig. 5.3A). However, the behavior of the global estimate $\langle\rho_{CH}\rangle$ first decreases until it reaches the order-disorder transition, then, in the ordered phase, it increases with increasing order (Fig. 5.3B). It becomes clear that the estimates differ qualitatively at the transition if their ratio $\langle\rho\rangle/\langle\rho_{CH}\rangle$ is plotted, which has a maximum at the transition (Fig. 5.3C). This means that the form of the shoal is at the transition badly approximated by a convex hull and/or that the density is very heterogeneous.

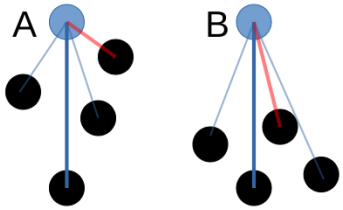

Figure 5.4: The 4th closest neighbor distance (blue line) is the same for the homogeneous (**A**) and heterogeneous (**B**) case in contrast to the nearest neighbor distance (red line).

A very heterogeneous density is badly estimated by the kth nearest neighbor density ρ. A more local estimate is necessary as for example the density estimate based on the nearest neighbor

$$\rho_{NND} = \frac{2}{\pi r_{NND}^2} \,. \tag{5.6}$$

An example where ρ_{NND} but not ρ distinguishes an homogeneous from a heterogeneous density is given in Fig. 5.4. The ρ_{NND} detects qualitatively the same minimum density at the transition as the global density estimate based on the convex hull (inset Fig. 5.3B). Therefore, the deviation from a convex form is not the only explanation. Also a strong heterogeneous density causes the density ratio $\langle\rho\rangle/\langle\rho_{CH}\rangle$ to peak at criticality.

Next to the heterogeneous distribution of the agents and non-convex form of the shoal, the transition is marked by a peak in the coefficient of variation of area of the convex hull (Fig. 5.3C inset)

$$CV(\text{Area}_{CH}) = \frac{\text{STD}(\text{Area}_{CH})}{\langle\text{Area}_{CH}\rangle} \,. \tag{5.7}$$

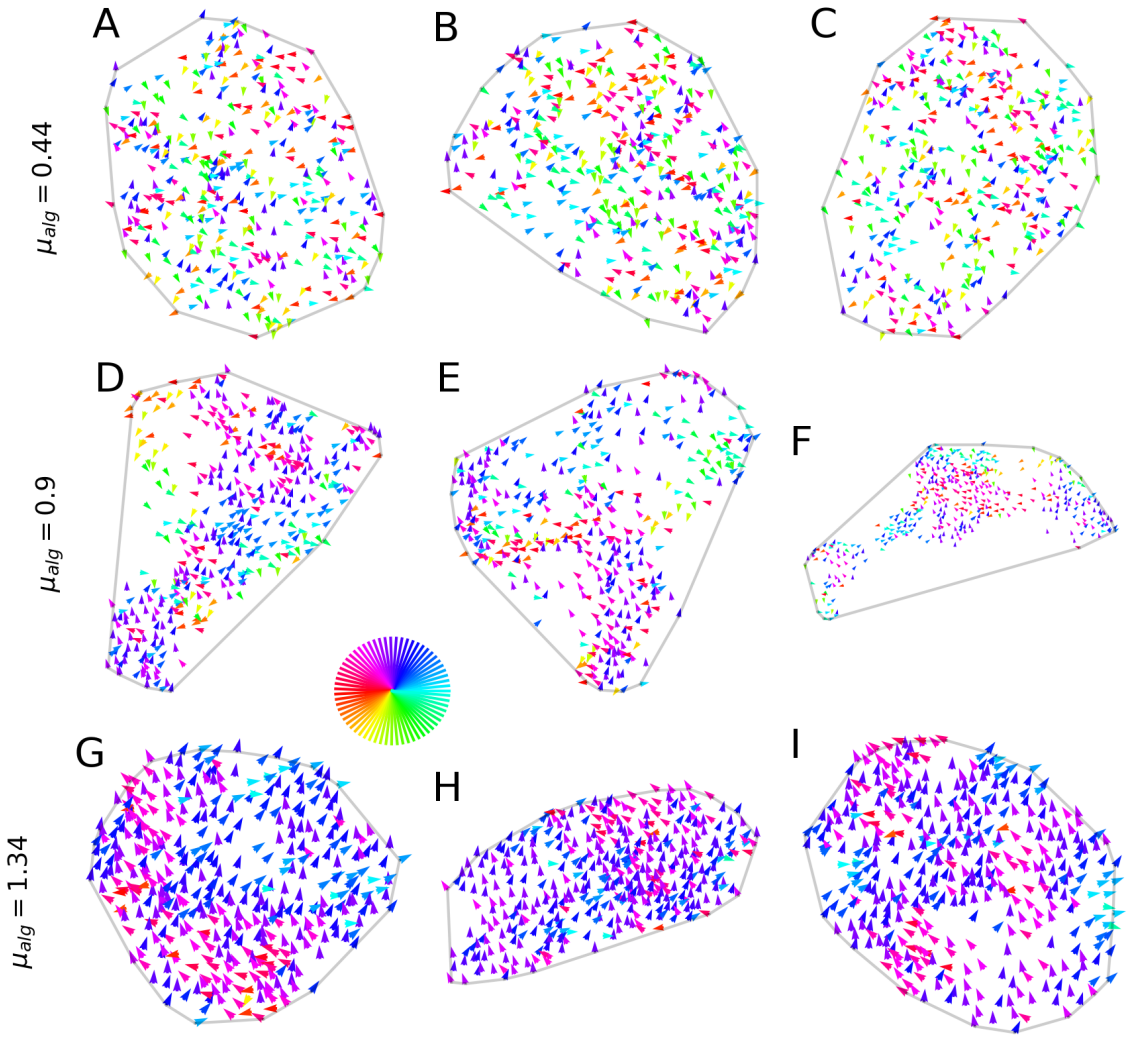

Figure 5.5: Snapshots of shoals at different alignment strength. 1st row (**A-C**): disordered phase ($\mu_{alg} = 0.44$). 2nd row (**D-F**): close to order-disorder transition ($\mu_{alg} = 0.9$). 3rd row (**G-I**): ordered phase ($\mu_{alg} = 1.34$). Each arrow represents an agent with its color as direction (indicated by the color-ball). The transparent line spanning each shoal is the convex-hull. All other parameters are listed in Tab. 3.1 (for clarity $D = 0.5$).

In Fig. 5.5 three independently initiated shoals are shown for three different parameter values. For the disordered (Fig. 5.5A-C) and ordered (Fig. 5.5G-I) shoals the convex hull approximates the swarm border well. Not so at the order-disorder transition (Fig. 5.5D-F, especially F). The samples also support the maximum of $CV(\text{Area}_{CH})$ at the phase transition (compare Fig. 5.5D and F). In addition, the heterogeneous agent distribution is well visible (compare Fig. 5.5E with B and H). Note, that the samples are randomly selected but rotated such that the groups heading direction is parallel to the y-axis.

The cause of the density anomalies is the balance between directional noise D and directional coupling strength μ_{alg} at the transition. In addition, it is linked to the directional correlation length, which diverges at the transition (in the thermodynamic limit). Local perturbations in the direction have about the same influence as the social information. These perturbations can cause the formation of small or large subgroups heading in the newly initiated, random direction. This process results in coherent subgroups that differ in their respective heading direction and therefore cause these complex structures at the transition. Note that this explanation is in agreement with a diverging correlation length at the transition (for finite systems: a correlation length growing with system size). The correlation strength decreases with distance for any point in the parameter space arbitrary close to (but not at) the transition. Furthermore, the correlation length is measured via fluctuations. The mere existence of fluctuations combined with a decay in correlation strength with distance is consistent with coherent local subgroups at the transition. Sure, the size of the subgroups can, but not necessarily must, span the whole shoal.

5.1.3 Shoal elongation and obliqueness

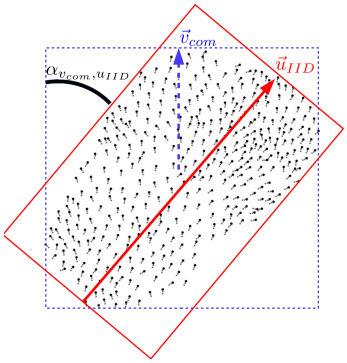

Figure 5.6: Minimal bounding boxes parallel to group velocity $\vec{v}_{com}$ (blue dashed box) and to the direction of largest swarm extend $\vec{u}_{IID}$(red box).

The shoal form and its orientation with respect to the group velocity are explored. I compare the elongation $AR_{v_{com}}$, i.e. the aspect ratio of the minimal bounding box oriented parallel to the group velocity $\vec{v}_{com}$ (Kolpas et al., 2013), and the aspect ratio $AR_{u_{IID}}$ of the minimal bounding box parallel to the vector connecting the agents with the largest inter individual distance $\vec{u}_{IID}$. An illustration of both minimal bounding boxes is shown in Fig. 5.6. The aspect ratio with respect to $\vec{u}_{IID}$ contains only more information and allows interpretation if the angle $\alpha_{v_{com},u_{IID}}$, also referred to as obliqueness, between group velocity and $\vec{u}_{IID}$ is also known.

In Fig. 5.7 the aspect ratios are plotted and equally distributed samples of the corresponding minimal bounding boxes. It is apparent that the combination of $AR_{u_{IID}}$ and $\alpha_{v_{com},u_{IID}}$ allows a much richer interpretation of the shape of the shoal compared to $AR_{v_{com}}$ (compare bounding boxes of Fig. 5.7a and b). While the $AR_{v_{com}}$ only decreases above the order-disorder transition and reaches its minimum in the ordered phase (Fig. 5.7a), the $AR_{u_{IID}}$ shows an increase already in the disordered phase close to the transition and peaks around the transition (Fig. 5.7b).

The angle $\alpha_{v_{com},u_{IID}}$ explains the difference between the aspect ratios. In the disordered phase, the obliqueness is random $\alpha_{v_{com},u_{IID}} = \pi/4$ and remains so until the phase transition (Fig. 5.7c). In the ordered phase, the obliqueness is $\alpha_{v_{com},u_{IID}} = \pi$, i.e. $AR_{u_{IID}} = (AR_{v_{com}})^{-1}$. Interestingly, there is another qualitative change for the low-noise and strong-alignment limit where the shoal has a vanishing obliqueness, i.e. where $AR_{v_{com}} = AR_{u_{IID}}$.

In summary, at the phase transition the obliqueness corresponds to the random value $\alpha_{v_{com},u_{IID}} = \pi/4$, typical for the disordered phase, and is combined with the transition typical large aspect ratio $AR_{u_{IID}}$.

For the later interpretation of the group response, the structure defined by density (Sect. 5.1.2) is more relevant than aspect ratio and obliqueness. Still, it is interesting, that in this model obliqueness is a dominant feature close to a phase transition. Also Kolpas et al. (2013) reported obliqueness of shoals in a three-zone model with Voronoi interactions. However, they did not report

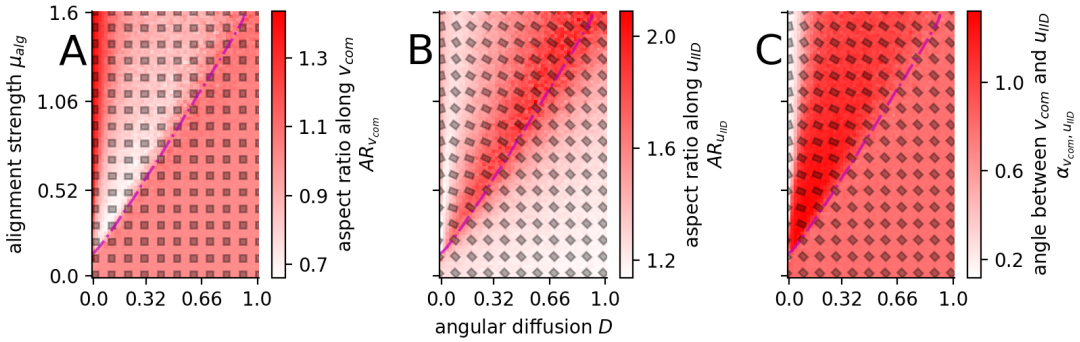

Figure 5.7: Shoal aspect ratios with respect to different directions. The aspect ratio of the minimal bounding box parallel to the group velocity $AR_{v_{com}}$ (**a**) and The of the box parallel to the vector $\vec{u}_{IID}$ connecting the two most separated agents $AR_{u_{IID}}$ (**b**). The angle $\alpha_{v_{com},u_{IID}}$ (**c**) between the $\vec{v}_{com}$ and $\vec{u}_{IID}$ allows an illustration of the minimal bounding boxes of $AR_{u_{IID}}$ (transparent boxes in **b,c**). All panels: bounding boxes are normalized with respect to their area and it is assumed that shoals move in positive y-direction.

obliqueness for metric interactions. Thus, it is likely that the topology is the cause of this feature. For example are the agents at the shoal edge connected to other edge-agents at a larger distance compared to bulk agents. These long-range connections could amplify initially local perturbations across the edge.

5.2 Response to fluctuations

The main motivation for the "criticality hypothesis" in group living animals is based on the group performance, because the quantities which diverge at a second order phase transition in statistical physics system are global averages. The most supporting quantity, the susceptibility (as discussed in Sect. 4.1.2), is the sum over all pair correlations of velocity-fluctuations, i.e. it is a group measure. I focus in this section on the "classical" physics quantities of the system (susceptibility, information transfer) and show that they indeed peak, as expected, at the transition.

The group measure susceptibility

$$\chi = N[\langle \Phi^2 \rangle - \langle \Phi \rangle^2] \tag{5.8}$$

is based on fluctuations of the order parameter Φ and the local information transfer

$$C(\delta\vec{v}_i, \delta\vec{v}_j) = \sum_{i,j \in \mathbb{N}_i} \delta\vec{v}_i \cdot \delta\vec{v}_j \tag{5.9}$$

on the correlation of the velocity fluctuations $\delta\vec{v}$ between interacting agent pairs. Both measures have been introduced and theoretically discussed in detail (Sect. 4). To recapitulate, χ can also be computed from the correlation of the velocity fluctuations over all possible agent pairs (see Eq. 4.14). Therefore, $C(\delta\vec{v}_i, \delta\vec{v}_j)$ represents a local version of the susceptibility, which I frame as *information transfer* but should be interpreted with caution (see Sect. 4.2.2).

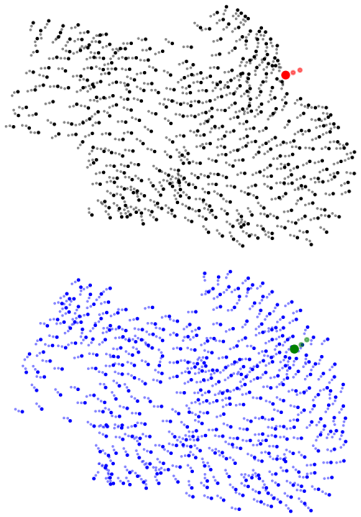

Figure 5.8: Responding (black dots) and non-responding prey (blue dots).

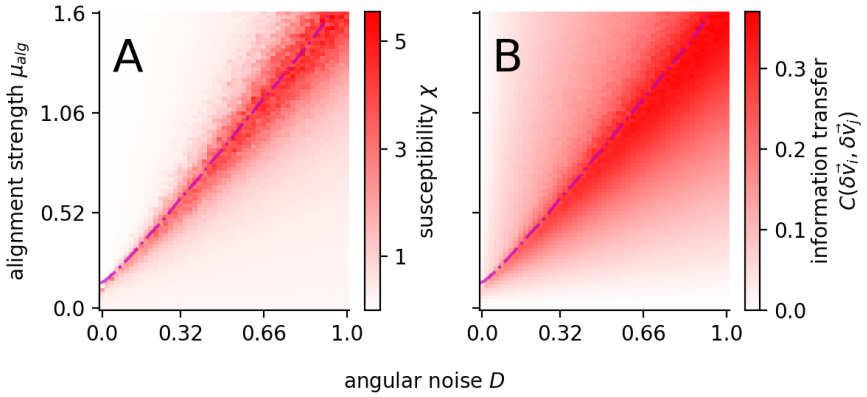

Figure 5.9: Susceptibility and local information transfer. **A**: susceptibility χ has been computed via the fluctuation of the order parameter Φ (Eq. 4.13) **B**: local information transfer is the correlation of the velocity fluctuations between interacting agents $C(\delta\vec{v}_i, \delta\vec{v}_j)$. Both panels: the dash dotted magenta line marks the order-disorder transition.

These measures are known to peak at the critical point (continuous phase transition) for models in statistical physics (Christensen and Moloney, 2005). From a functional point of view, they make the critical point a potentially preferable region of parameter space.

As shown in Fig. 5.9, also in our spatially explicit schooling prey model both measures peak at the phase transition. Note that both measures approximate the response to a small, global perturbation (see Sects. 4.1.4, 4.2.2). This makes the measure general, i.e. they are independent of the predator stimulus, but also very specific, because it measures response based on steady states and small forces.

5.3 Response to a predator

This section simulates the attack of a predator (described in Sect. 3.2) and how the shoal and its biologically relevant parts avoids the predator by a simple repulsion mechanism. I quantify this problem specific response by classical distance measures and by measures that carry a biological meaning. The latter allow an interpretation with respect to group-level selection. Social prey that is unresponsive to the predator will serve as a crucial control for enhanced avoidance due to shoal structure.

5.3.1 Control for structure: non-fleeing prey

I have already highlighted in Sect. 5.1.2 that the structure of the shoal does change at the phase transition. This density variation is intrinsically linked with the order-disorder transition and happens even if the parameters that obviously influence the density, i.e. the preferred distance r_0 and the steepness of the distance regulating force m_d, remain constant.

To control against these structural changes, I introduce the concept of non-fleeing (**NF**) prey, which are prey agents identical in parameters, position and velocity to the normal prey but who do not flee from the predator. The NF-prey is an instantaneous copy of the normal prey at the time the predator appears. However, after their initialization they evolve independently as illustrated in Fig. 5.8 for a short time after initialization. Also an additional predator is created, again with the same position and velocity as the original predator. This predator copy behaves exactly as the original predator but only senses the NF-prey.

Therefore, the NF-prey is not only a control against the structure of the shoal but also against the default behavior of the predator. Since the predator adapts its direction following the weighted average of its frontal nearest neighbors, the predator will move different if it follows NF-prey.

However, I will mostly refer to the NF-prey as a control against the shoal structure (but it is actually more).

5.3.2 Distance to the predator

An obvious and simple measure of how well the shoal performs is the average distance to the predator

$$\langle r_{ip} \rangle = \langle |\vec{r}_i - \vec{r}_p| \rangle \ . \tag{5.10}$$

However, from a biological perspective the distance to the predator is for agents located close and in front of the predator more relevant than for example agents behind the predator. These close and frontal agents are potential targets of the predator and I refer to their average distance to the predator as $\langle r_{ip} \rangle_{knn}$ (with $k = 10$). For an illustration see Fig. 5.10.

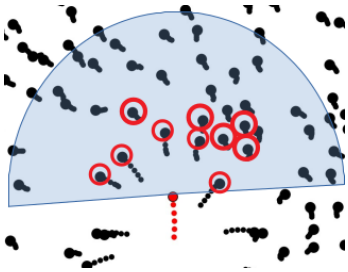

Figure 5.10: Predator's frontal K-nearest neighbors.

Both measures have a maximum at the order disorder transition (Fig. 5.11A, B). Note that the maximum of the biological relevant distance $\langle r_{ip} \rangle_{knn}$ is not as pronounced as the predator distance averaged over the whole shoal. This indicates that not the response but the shoal structure is causing the maximum, which is plausible because of the reported minimum density at the transition (Sect. 5.1.2).

By computing the same measures for the NF-prey (inset in Fig. 5.11A, B) and subtracting them from the already reported measures, the maximum vanishes for both structure-controlled measures (Fig. 5.11C, D). It also reveals that only in the ordered phase the prey increase their distance to the predator compared to NF-prey.

5.3.3 Capture rate

By estimating the predator distance to the frontal k-nearest neighbors, I already introduced a biological meaningful measure. However, a real predator consumes its prey and even if a closer prey is more likely to be caught (as also implemented in the model), to limit oneself to a mere distance analysis is incomplete in the context of predator-prey dynamics. Therefore, I computed the predator capture rate γ_c as number of prey captured per time unit. The lowest capture rate is detected at the transition; however, there the non-fleeing prey are as well caught at the lowest rate (Fig. 5.12A), suggesting that the shoal structure is causing the minimum.

I control against the structure by subtracting the capture rate of non-fleeing prey $\mathrm{NF}(\gamma_c)$. Indeed, the structure controlled capture rate has no minimum at the transition (inset Fig. 5.12B). However, the difference in capture rate is not unambiguous as explained in the following example case. If in two different parameter regions P_a and P_b the same difference is detected $\gamma_c - \mathrm{NF}(\gamma_c) = -0.5$, it does not mean that in both cases the prey avoids equally well the predator. If the capture rate is $\gamma_c = 0$ in region P_a and $\gamma_c = 1$ in region P_b , all prey could escape in region P_a but only half of the prey escaped in region P_b.

A measure that detects the difference between region P_a and P_b is the escape ratio, i.e. the fraction of surviving fleeing prey, which would have been captured if they would not flee

$$R_{esc} = 1 - \frac{\gamma_c}{\mathrm{NF}(\gamma_c)} \ . \tag{5.11}$$

It is equal to the difference between the capture rates of non-fleeing and fleeing prey $\mathrm{NF}(\gamma_c) - \gamma_c$ scaled by $\mathrm{NF}(\gamma_c)$.

The escape ratio R_{esc} has no minimum at the phase transition but rather gradually increases with order (Fig. 5.12B).

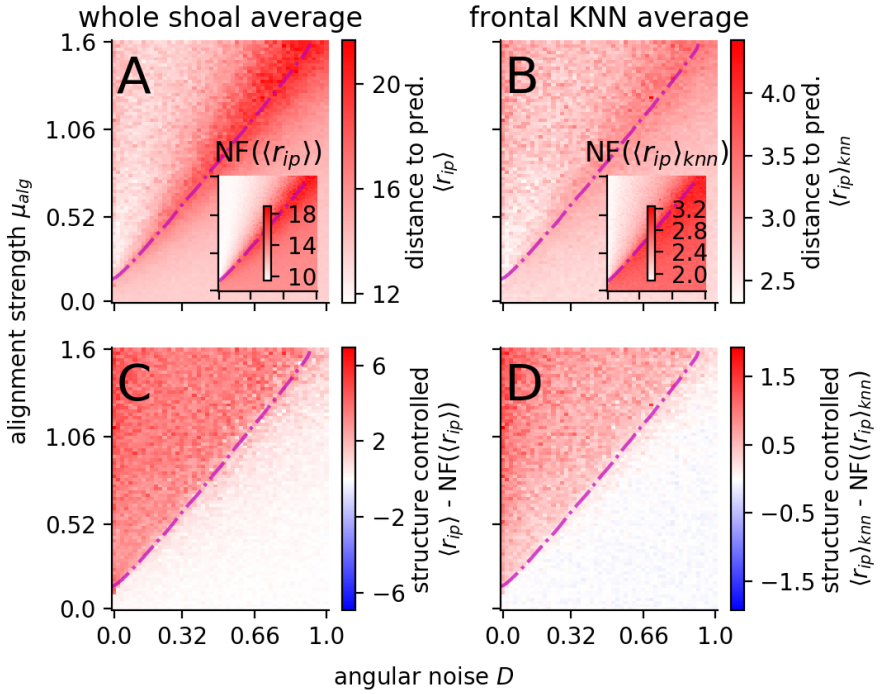

Figure 5.11: Distance to the predator with and without control against structure. The distance to the predator r_{ip} is averaged over the whole shoal $\langle r_{ip}\rangle$ (**A**) and only over the frontal k-nearest neighbors $\langle r_{ip}\rangle_{knn}$ (**B**) with $k = 10$. To control against the shoal structure, the same measures are computed for the non-fleeing (NF) prey (insets in **A, B**). The structure-controlled distance to the predator is computed for the shoal average $\langle r_{ip}\rangle - \mathrm{NF}(\langle r_{ip}\rangle)$ (**C**) and the k-nearest neighbor average $\langle r_{ip}\rangle_{knn} - \mathrm{NF}(\langle r_{ip}\rangle_{knn})$ (**D**).

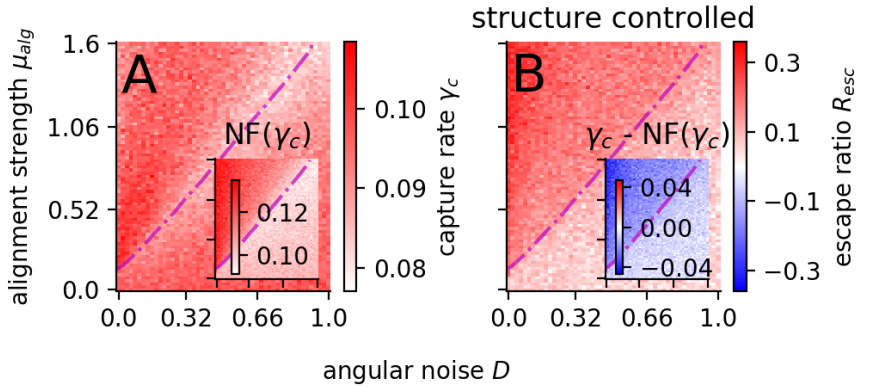

Figure 5.12: Predator's capture rate and preys escape ratio. **A**: predator's capture rate for fleeing γ_c and non-fleeing NF(γ_c) prey (inset). **B**: the escape ratio $R_{esc} = 1 - \gamma_c/\mathrm{NF}(\gamma_c)$ and the difference in capture rates between fleeing and non-fleeing prey (inset).

5.4 Robustness: parameter dependence

A detailed robustness analysis for the changes in the parameters of the preys and predator and the dependence on the predator attack schemes can be found in the appendix in Sect. A.4. It supports the main findings: the group optimum is at criticality but its cause are structural features rather than prey responsiveness. A noteworthy finding is that a wall effect exists for a specific predator attack (straight path through the swarm with no angular variation in initiation position). That means the fleeing prey build up a dense wall in front of the predator (see Sect. A.4.3).

An crucial results of the robustness analysis is that the signal strength shifts the group optimum away from the critical point into the ordered regime. Therefore, this is presented below, for the remaining robustness analysis I refer to Sect. A.4.

5.4.1 Signal strength dependence

To frame the flee strength μ_{flee} as the strength of the signal helps to interpret this variation. If I connect the signal strength to the considerations of Sect. 4.1.4, I am now able to see how the violation of the small force assumption acts in combination with a local force. Recall, the link of the susceptibility to the fluctuations in the order parameter is valid for differential changes of the external field.

In the section before a possible responsive maximum in R_{esc} was located for the default flee strength ($\mu_{flee} = 4$) at an alignment strength of $\mu_{alg} \approx 2$. In this larger parameter scan the maximum of the escape ratio is well visible (Fig. 5.13B). Most importantly, the maximum increases and shifts to larger alignment strength with increasing signal strength μ_{flee}.

While the responsive optimum is for the default flee strength only detectable via the structure-controlled measures, for larger flee strength a local maximum appears also for structure uncontrolled measures as the distance to the predator $\langle r_{ip} \rangle$ and the capture rate γ_c (Fig. 5.13C-E). For sufficient strong signals this second peak is even the global maximum (Fig. 5.13D,E).

To ensure the correct interpretation of this second peak I compute the total number of prey ever detected by the predator, i.e. the number of unique frontal nearest neighbors of the predator. Indeed, also this measure shows a distinct minimum close to the responsive optimum, and for strong signal even global, group optimum (Fig. 5.13F).

5.5 Summary

The focus of this chapter was the analysis of the group-performance of the collective in the presence of a predator. The main motivation for the "criticality hypothesis" in group living animals is based on the group performance, because the quantities which diverge at a second order phase transition in systems from statistical physics are global averages. The most supporting quantity, the susceptibility, is the sum over all velocity-fluctuation pair correlations, i.e. it is a group measure. Since the system under consideration is a non-equilibrium system and the signal is a non-standard, in statistical physics terms, perturbation, I showed that the "classical" physics quantities of the system (susceptibility, information transfer) indeed peak at the transition. In addition, the response to a local, inhomogeneous perturbation, i.e. the predator, is best at the transition. This is also supported by more biological relevant measures, as the capture rate or the distance to possible targets. However, when the results are corrected against structural effects (with a non-fleeing control shoal), the "responsive group optimum" shifts into higher order regions, away from the phase transition. This suggests that the group structure is causing the group optimum at the transition. It is supported by the strong interplay between spatial group structure and the order-disorder transition (lowest density at transition). The robustness analysis revealed that for stronger flee strengths the group optimum shifts to the ordered region, i.e. away from the phase transition. This is even true for not spatially controlled response measures.

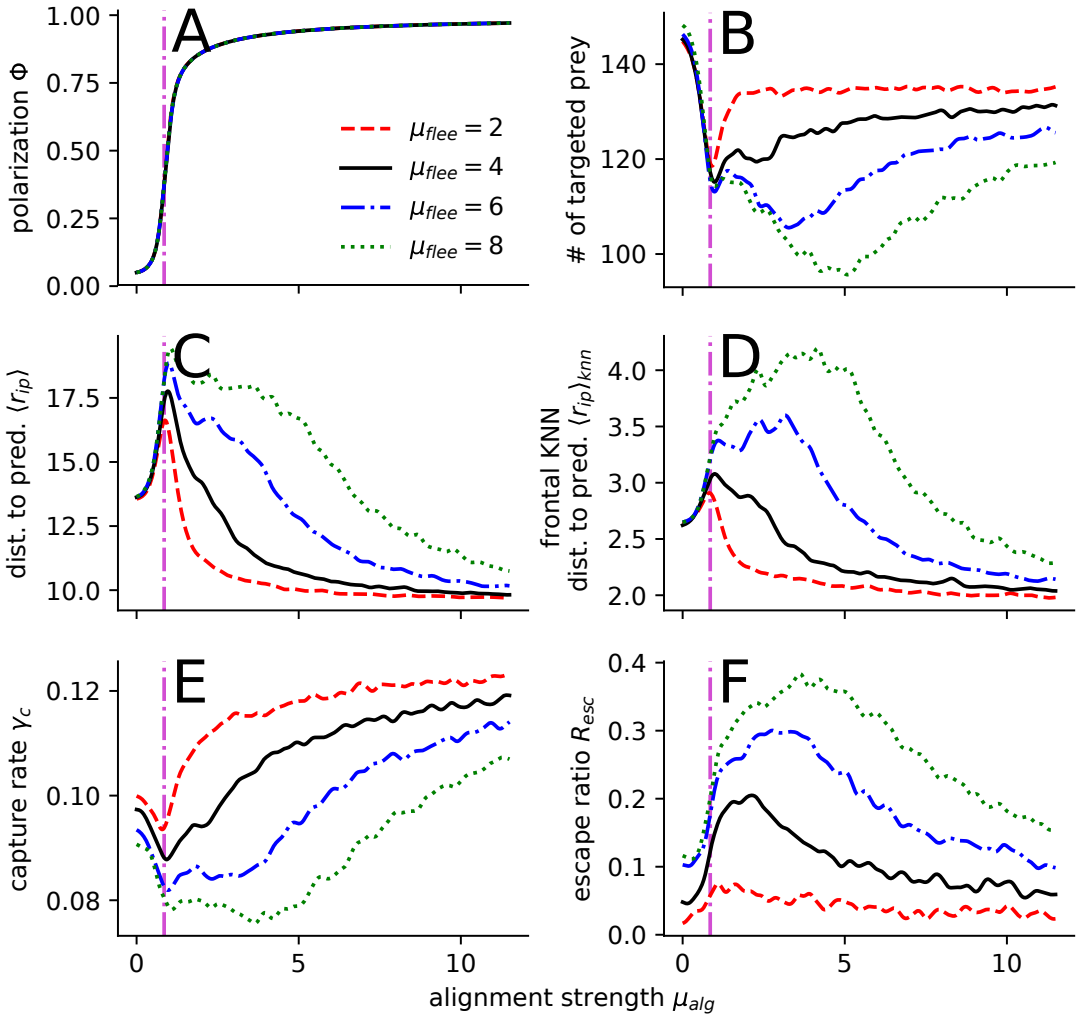

Figure 5.13: Signal strength dependence of group measures. The polarization Φ (**A**), total number of targeted prey (**B**), distance to predator averaged over the whole shoal $\langle r_{ip} \rangle$ (**C**) and only over the frontal k-nearest neighbors $\langle r_{ip} \rangle_{knn}$ (**D**), capture rate γ_c (**E**) and the escape ratio R_{esc} (**F**) are computed for a wide range of alignment strength and for different flee strengths μ_{flee}. The line styles of the flee strength $\mu_{flee} \in [2,\ 4,\ 6,\ 8]$ are indicated in the legend of panel **A**.

6 Evolution under predation

In this chapter, I will finally test the criticality hypothesis in a collective under predation in an evolutionary context. An individual prey could in principle evolve in 5 parameters. However, because the distance to the order-disorder transition is of greatest importance for the criticality hypothesis, the prey only evolves in its individual alignment strength μ_{alg} in this study. The chapter is structured as follows. First, the evolutionary algorithm and measures necessary to analyze the evolution are introduced. Second, the evolution and the corresponding fitness gradient of the system are analyzed. There are two surprising and important results which I will analyze in depth, (i) the order-disorder transition is an "evolutionary accelerator", i.e. an evolutionary unstable point where selection gradients become maximal. It leads to a strong evolutionary driving away from the transition region, which pushes the system out of the transition region. And (ii) the evolution equilibrates in the strongly order phase. Third, the robustness of the findings is checked by varying aspects of the prey, predator and of the selection mechanism. Forth, I try to tune the evolutionary stable state (ESS) to the phase transition by an additional selection mechanism.

6.1 Evolutionary algorithm and measures

Here, I introduce the evolutionary algorithm and measures that are important to analyze the evolution. The latter are the estimation of the fitness gradient, the evolutionary stable state and the spatial self-sorting of individuals in the shoal.

6.1.1 Evolutionary algorithm

The evolutionary algorithm is designed to mimic a simplified natural selection at the level of behavioral phenotype. Among others, the influence of fecundity selection or sexual selection is neglected. The fitness function is only based on how likely an individual is captured in a predator attack. It is a biologically reasonable simplification in the context of predator-prey interactions. The algorithm consists of (i) a fitness estimation step, (ii) a fitness-proportionate-selection step and (iii) a mutation step.

(i) The fitness is estimated by running $N_f = 76$ independent attack simulations on the same phenotype population. For each simulation the $\gamma_a \cdot T_{simu}$ agents with the highest cumulative probability of being caught (Eq. 3.13) are declared as dead. The fitness of agent i is:

$$f_i = -N_{c,i} + max(N_{c,j}, j) \ . \qquad (6.1)$$

Here $N_{c,i}$ is the number of simulations in which agent i was captured and $max(N_{c,j}, j)$ is the largest number of deaths among all agents.

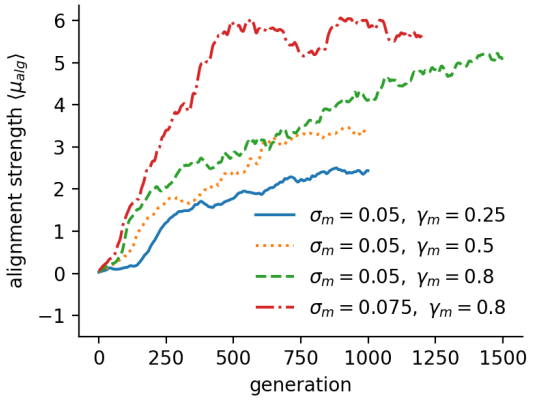

Figure 6.1: Evolution under different mutation rate γ_m and strength σ_m. Note that the predator only goes straight in the direction of the COM at initiation and keeps this direction. Therefore, the evolutionary stable state does differ from the standard predator simulations.

(ii) The N offspring are generated via the fitness-proportionate-selection. Thereby, an offspring

inherits the parameters of the parent i with probability

$$p_{parent,i} = \frac{f_i}{\sum_j f_j} \ . \tag{6.2}$$

(iii) An offspring agent mutates with a probability γ_m, the mutation rate, by adding to its alignment strength μ_{alg} a Gaussian random variable with zero mean and standard deviation σ_m, as the mutation strength.

Steps (i) until (iii) are repeated in each generation.

Note that instead of step (i) the agents could directly be captured during the simulation and would be removed from the group during the run. However, this introduces an additional source of noise in the predation process and the resulting fitness gradient of the prey would become more noisy. As a consequence the number of generations needed to reach an ESS increases. A version of this alternative is later explored in Sect. A.6.3.

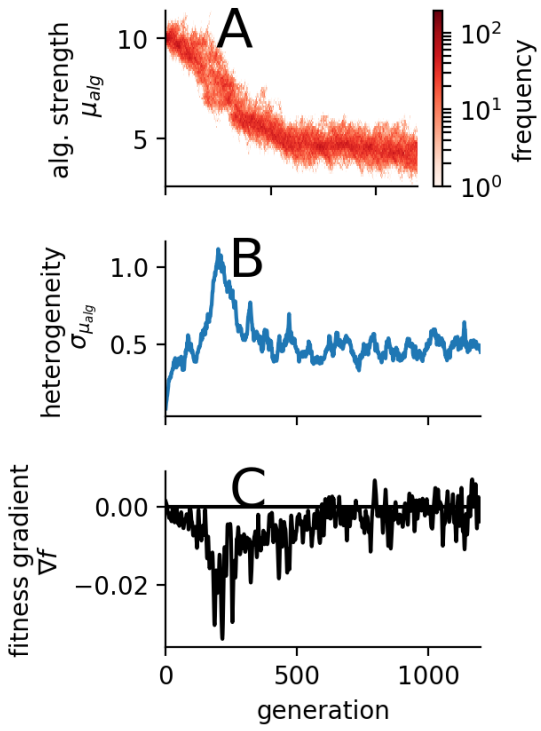

Figure 6.2: In an evolution run with a prey population initiated at $\mu_{alg} = 10$ over 1200 generations (**A**) does the heterogeneity first increases and then fluctuates around the a stable value of $\sigma_{\mu_{alg}} \approx 0.5$ (**B**). The estimated fitness gradient ∇f (**C**) correlates with the heterogeneity.

The parameters for the mutation rate and strength are $\gamma_m = 0.8$ and $\sigma_m = 0.075$, respectively. With these parameters the evolutionary stable state is reached in about 500 generations, which is computationally manageable, and it keeps the fluctuations around the ESS low (Fig. 6.1). For slower rates and strength the ESS is not reached, even after 1500 generations.

6.1.2 Estimation of the fitness gradient and ESS

In the evolutionary algorithm the finite mutation strength and the stochastic roulette-wheel selection introduce noise on top of the intrinsic stochasticity of the predator-prey dynamics (Eq. 3.1). This stochasticity is essential for evolutionary adaptation and exploration of the phenotype space, but makes it challenging to identify the evolutionary stable states (ESS) with high precision in evolutionary simulations.

To circumvent this uncertainty about the exact optimum I estimate the evolutionary stable state based on the zero-crossing of the fitness-gradient estimated from numerical simulations. For a population in generation g with agent parameters $\vec{\mu}_{alg}(g) \in \mathbb{R}_+^N$ the estimated fitness gradient $\nabla f(g)$ is computed by predicting the mean outcome of the fitness-proportionate selection

$$\langle \mu_{alg} \rangle_{predict}(g) = \vec{p}_{parent,i} \cdot \vec{\mu}_{alg} \tag{6.3a}$$

$$= \frac{1}{\sum_j^N f_j} \sum_i^N f_i \mu_{alg,i} \tag{6.3b}$$

and subtracting from it the current mean-value:

$$\nabla f(g) = \langle \mu_{alg} \rangle_{predict} - \langle \mu_{alg} \rangle \ . \tag{6.4}$$

Note that, in sake of readability, I omitted for terms on the RHS of Eqs. 6.3, 6.4 the dependency on the generation g.

The average fitness gradient at a specific alignment strength is

$$\nabla f(\mu_{alg}, \Delta_\mu) = \langle \nabla f \rangle_{\mathbb{S}_{\mu_{alg},\Delta_\mu}} = \frac{\sum_{g \in \mathbb{S}_{\mu_{alg},\Delta_\mu}} \nabla f(g)}{|\mathbb{S}_{\mu_{alg},\Delta_\mu}|} \tag{6.5}$$

where $\mathbb{S}_{\mu_{alg},\Delta_\mu}$ is the set of generations that fulfill the condition:

$$\mu_{alg} - \Delta_\mu/2 \leq \langle \mu_{alg} \rangle(g) \leq \mu_{alg} - \Delta_\mu/2 \ . \tag{6.6}$$

Therefore, Eq. 6.5 represents a simple binning of generations with a bin-width of Δ_μ. The maximum of the estimated fitness landscape, i.e. the evolutionary stable state, is where the estimated fitness gradient is zero and where its slope is negative.

Influence of population heterogeneity on fitness gradient.

The implemented evolution algorithm does intrinsically regulate the heterogeneity of the population $\sigma_{\mu_{alg}}$. This regulation is due to a negative feedback between $\sigma_{\mu_{alg}}$ and the fitness gradient ∇f. In a strongly heterogeneous population the individual behavior differs strongly between the agents. This causes strong relative fitness differences and therefore increases ∇f. If the fitness gradient is large, the roulette wheel algorithm will create a more homogeneous offspring population. Thus, it down regulates the heterogeneity. Still the process is stochastic which can cause strong variations in heterogeneity. Therefore, one should always keep in mind that large fitness gradients can be due to random fluctuations in heterogeneity. For example, consider the evolution presented in Fig. 6.2A where the ESS is reached after about 600 generations. One can clearly see that the fitness gradient is more correlated to the heterogeneity (Fig. 6.2B, C) than to the distance to the ESS. In fact the correlation is $C(\nabla f, \sigma_{\mu_{alg}}) \approx -0.76$ for the generations before equilibration, i.e. before the 600th generation.

6.1.3 Self-sorting

In heterogeneous systems where agents are able to change their relative position, self-sorting is expected to happen (Couzin et al., 2002). A simple examples of self-sorting in cohesive collectives is that faster agents sort to the front (Couzin et al., 2002). I quantify self-sorting via the Pearson correlation coefficient between the alignment parameter $\mu_{i,alg}$ of individual agents and their mean relative location within the collective $\langle r_{i,x} \rangle$ where $x \in \{f, s, d\}$ which stands for front, side and local density respectively

$$C(\mu_{alg}, x) = \frac{(\mu_{alg} - \langle \mu_{alg} \rangle)(x - \langle x \rangle)}{\sigma_{\mu_{alg}} \sigma_x} \ . \tag{6.7}$$

Agents at the front (back) have the largest (smallest) front-location and at the side (center) have the largest (smallest) side-location. The local density sorting is the correlation of the agents local density (see Eq. 5.3 in Sect. 5.1.2) and its alignment strength. Another, more general, quantification of self-sorting is how associative the agents sort/mix. The implementation of the assortativity coefficient (Newman, 2002) in igraph is used on the interaction network (Voronoi) with the values for each agent corresponding to their alignment strength. A detailed definition on the relative position and assortativity is given below.

Relative positions

In order to define the relative positions with respect to the front and to the side, first every agent-position is represented by its distance to the center of mass of the collective

$$r_{i,com} = |\vec{r}_{i,com}| = \vec{r}_i - \vec{r}_{com} \quad \text{with } \vec{r}_{com} = \sum_i \vec{r}_i / N \tag{6.8}$$

and the angle between its position and the mean velocity of the collective

$$\alpha_{i,com} = \angle(\vec{r}_{i,com}, \vec{v}_{com}) \quad \text{with } \vec{v}_{com} = \sum_i \vec{v}_i \ . \tag{6.9}$$

I refer to this representation as the *folded polar swarm-coordinates*.

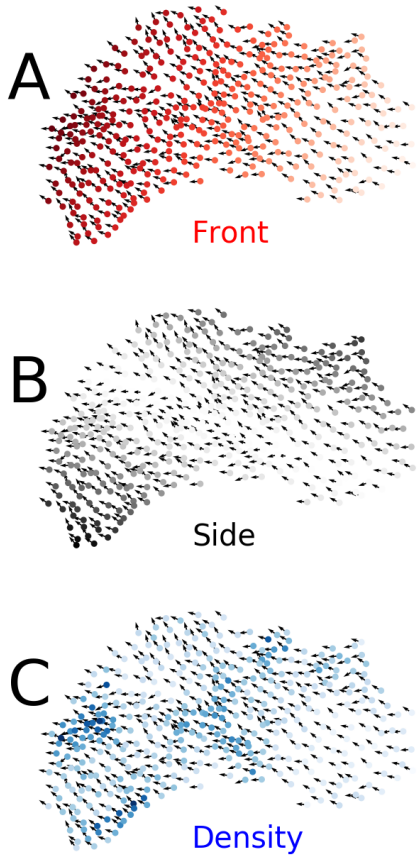

Figure 6.3: Relative shoal position and density. Color coded spatial position for the shoal's front (**A**) side (**B**) and densest (**C**) regions.

Note that the x-axis is parallel to $\vec{v}_{com}$, the center of mass is at the origin and the quadrants IV and III are folded onto I and II respectively. The folding is reasonable because of the left-right symmetry. The relative front position is

$$\tilde{r}_{i,f} = r_{i,com} \cos \alpha_{i,com} \tag{6.10}$$

and it is normalized via

$$r_{i,f} = \frac{\tilde{r}_{i,f} - \min(\tilde{r}_{j,f}, j)}{\max(\tilde{r}_{j,f}, j) - \min(\tilde{r}_{j,f}, j)} \ . \tag{6.11}$$

It results in front positions in the interval $r_{i,f} \in [0, 1]$ (Fig. 6.3A).

The relative side-position is

$$\tilde{r}_{i,s} = r_{i,com} \sin \alpha_{i,com} \tag{6.12}$$

and the normalized version computes to

$$r_{i,s} = \tilde{r}_{i,s} / \max(\tilde{r}_{j,s}, j) \ . \tag{6.13}$$

Relative side positions are illustrated in Fig. 6.3B.

The normalization of both measures allows to compute a less noisy mean relative position over independent initializations. This is helpful because I am interested in weather an individual is at the front, not not how far the front is from the center of mass. Thus, the average normalized relative-position over S samples is

$$\langle r_{i,x} \rangle = \frac{\sum_{k=1}^{S} r_{i,x,k}}{S} \tag{6.14}$$

with $r_{i,x,k}$ as the normalized relative position of agent i in the kth sample run.

An analog normalization, as done for the front position (Eq. 6.11), and averaging is applied for the local density estimate of an individual agent (Fig. 6.3C).

Assortativity

The assortativity r is defined as

$$r = \frac{1}{\sigma_q^2} \sum_{j,k} jk(e_{j,k} - q_j q_k) \tag{6.15}$$

with $e_{i,j}$ as the joint probability that a randomly drawn edge connects vertices of type i and j, and q_x is the probability that a node of type x is at one end of a randomly drawn edge, i.e. it is the fraction of edges that have a vertex of type x at one end. The assortativity is the Pearson correlation coefficient over the values of the vertices connected by edges.

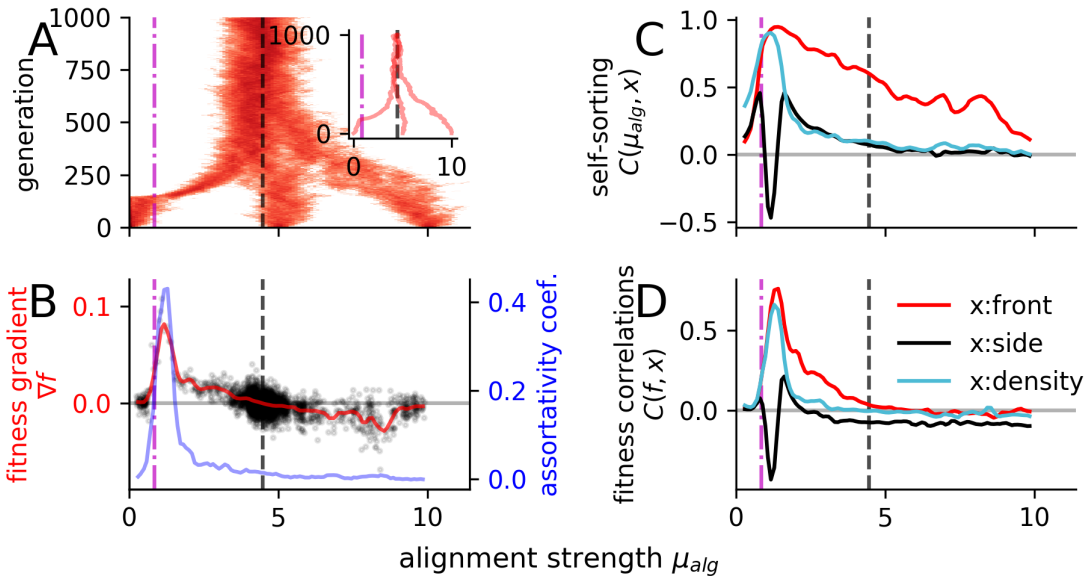

Figure 6.4: Evolution under predation. **A**: overlay of three evolutionary runs starting at $\langle \mu_{alg} \rangle = [0, 5, 10]$ over 1000 generations. The behavioral phenotype is determined only by the alignment strength as the evolving parameter. The predator attacks from random initial directions for $T_{simu} = 120$. The inset shows the evolution of the population mean alignment parameter $\langle \mu_{alg} \rangle$ of the three independently initialized runs. **B**: assortativity coefficient (blue line) and smoothed fitness gradient ∇f (red line). The evolutionary stable state is defined by the zero crossing of the fitness gradient and is represented by a vertical dashed black line. Black dots are the non-averaged fitness gradients for each generation. **C**: self-sorting measured as correlation $C(\mu_{alg}, x)$ between the individual alignment strength μ_{alg} and average spatial property of the individual as front-(red) and side-position(black) and local density (blue). **D**: correlation $C(f, x)$ of individual fitness with the, latter mentioned, average relative spatial positions. In all panels: the vertical dash-dotted magenta line marks the order-disorder transition and the vertical dashed black line the evolutionary stable state.

6.2 Evolution trajectory

The group-optimum at criticality with respect to prey-survival does not need to coincide with the evolutionary stable state (ESS) with respect to evolutionary adaptations on the individual level. To explore whether the transition region is favored by evolution, I let the individual alignment strength μ_{alg} evolve over 1000 generations, while keeping the angular noise constant ($D = 0.5$: center in Fig. 5.2). In this one-dimensional search-space the system can explore all collective states. I repeat the evolutionary simulations from different initial conditions: below ($\langle\mu_{alg}\rangle = 0$), above ($\langle\mu_{alg}\rangle = 5$) and far above ($\langle\mu_{alg}\rangle = 10$) the transition ($\mu_{c,alg} \approx 0.9$). To ensure that the evolution ends at the ESS, I compute the fitness gradient which represent the strength of the selection pressure at a specific mean alignment strength (see Sect. 6.1.2). Assuming a mono modal phenotype distribution, as observed in the evolutionary runs, a change in sign of the fitness gradient marks the location of the ESS. All three initiations end in the ordered region far above the critical point (Fig. 6.4A) and fluctuate around $ESS(\mu_{alg}) \approx 4.4$ (vertical dashed line Fig. 6.4B). Thus, the transition region is not an attractor of the evolutionary dynamics. On the contrary, it is a highly unstable point with fast evolutionary dynamics due to particularly strong selection pressure at criticality. The fitness gradient peaks shortly above the transition in the ordered phase (Fig. 6.4B), with evolutionary dynamics pushing the system out of the transition region towards stronger alignment.

In the following, I will analyze both findings, i.e. (i) the transition is an evolutionary accelerator and (ii) the ESS lies in the ordered region.

6.2.1 Order-disorder transition as an evolutionary accelerator

A possible driver of the maximal selection pressure at the transition is self-sorting, i.e. the tendency of individuals to sort according to their behavioral parameters along specific spatial dimensions of the school, e.g. front-back or side-center, or in regions of higher or lower density (Fig. 6.4C) (Couzin et al., 2002). I quantify this self-sorting through the Pearson correlation coefficient between the individuals alignment strength and its relative position/property (see Sect. 6.1.3). A more advanced measure for self-sorting is assortative mixing of the school (assortativity coefficient, Sect. 6.1.3). Assortativity (Fig. 6.4B) as well as the other self-sorting measures (Fig. 6.4C) exhibit extrema that coincide with the fitness gradient peak. Note that a strong assortative mixing is equivalent to the formation of spatially coherent subpopulation within the school exhibiting similar behavioral parameter. In this context, a peak in fitness gradient close to the transition means that subpopulations with stronger alignment, thus better directional coordination, actively or passively perform better at avoiding capture. An increase in the escape ratio R_{esc} with increasing alignment close to criticality (see Fig. 5.12B) suggest an enhanced active avoidance. However, passive effects appear to play an important role since the correlation between the fitness of a prey and its relative position becomes maximal in the same parameter region (Fig. 6.4D). One specific mechanism of passive avoidance is the dilution effect(Krause and Ruxton, 2002) caused by local density differences correlating with behavioral phenotypes. Stronger aligning individuals form denser subgroups within the prey school (density-sorting Fig. 6.4B). Consequently, they have systematically smaller domain of dangers (James et al., 2004) and are thus less frequently attacked by the predator.

Disentangle passive and active fitness benefits

It is possible to disentangle passive, structural effects from an active response, by setting the flee-strength to zero. This results in a significantly smaller, yet finite, fitness-gradient-peak at the transition (Fig. 6.5A). This suggests that both, the structural, passive selection as well as the different active avoidance behavior of different phenotypes contribute to the strong selection pressure at criticality. Interestingly, the fitness gradient seems to increase with the flee strength and saturates at $\mu_{flee} = 4$. However, $\mu_{flee} = 3$ has only about the same fitness gradient at the transition because the heterogeneity of the population $\sigma(\mu_{alg})$ is larger at the transition, which is likely due to stochasticity (Fig. 6.5B). Larger behavioral differences between individuals also

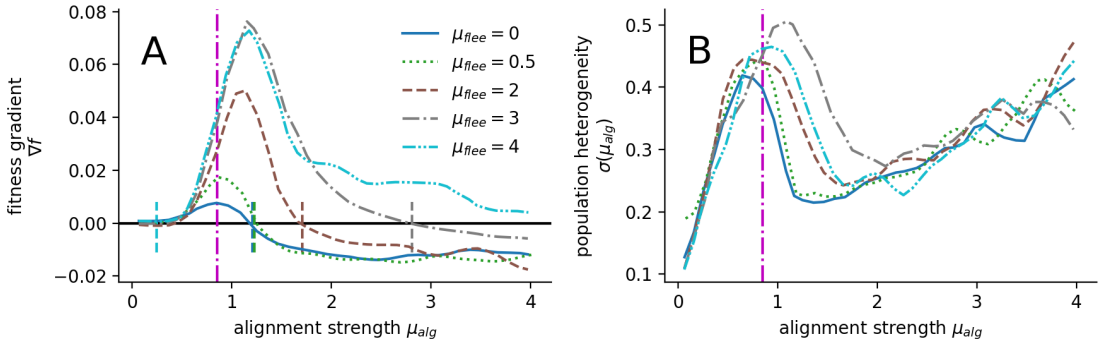

Figure 6.5: Fitness gradient and population heterogeneity for different flee strength. Mean fitness gradient ∇f (**A**) and population heterogeneity $\sigma(\mu_{alg})$ (**B**) estimated from evolution runs (similar to Fig. 6.4A) for different flee strength μ_{flee}. The line style and color represent the flee strength as shown in legend of panel **A**. The line styles of the flee strength $\mu_{flee} \in [0,\ 0.5,\ 2,\ 3,\ 4]$ are indicated in the legend of panel **A**. The magenta dash dotted vertical line marks the phase transition.

increases their relative fitness, causing a larger fitness gradient (see also Sect. 6.1.2).

The animation of the simulations reveal a likely explanation for the increase in the fitness gradient with increasing flee strength. The stronger the flee strength the more likely the predator will follow a fleeing prey for a longer time: if it does not flee at all the predator quickly catches up, overtakes and follows the next closest prey target group. On the other hand, if the prey avoids well, it stays longer in front of the predator and influences the predator's path. Additionally, the predator has a much lower relative velocity when it comes from behind the shoal and can therefore follow better. This combination leads to the effect, that the predator reaches less often the front of the shoal but always reaches the shoal end. This creates a fitness benefit for frontal individuals that increases with self-sorting and flee strength, as observed.

Self-sorting by symmetry breaking

Note that the sudden increase of self-sorting at the transition is due to a coupled symmetry breaking. At the order-disorder transition the directional symmetry is broken and the school "agrees" on a common movement direction. This also breaks the symmetry between relative locations within the school. For example, in the disordered phase, every edge position is equivalent. With the emergence of the common movement direction, the sides and rear of the school become distinguishable from the front. This can be clearly seen in the comparison of the correlations of individual alignment strength and specific relative spatial positions within the school ("side-sorting" versus "front-sorting"). Below the transition, the corresponding curves are indistinguishable, whereas at and above the transition they start to deviate and show different behavior with increasing alignment strength (Fig. 6.4C).

6.2.2 ESS dependence on signal strength

Despite the importance of self-sorting for the maximal selection pressure at the transition, it does not provide any explanation for the observed location of the ESS. The spatial properties can not explain the negative fitness gradient for $\mu_{alg} > ESS(\mu_{alg}) \approx 4.4$. In this regime either the self-sorting is negligible, as for side- and density-sorting (Fig. 6.4B), or the relative location has no effect on the individual fitness, as observed along the front-back sorting (Fig. 6.4E). If the ESS is not determined by the structural self-organization of the school, it has to originate from individually optimal predator evasion. In this case, the ESS has to depend on the flee-strength μ_{flee} as the main parameter tuning the strength of individual predator response.

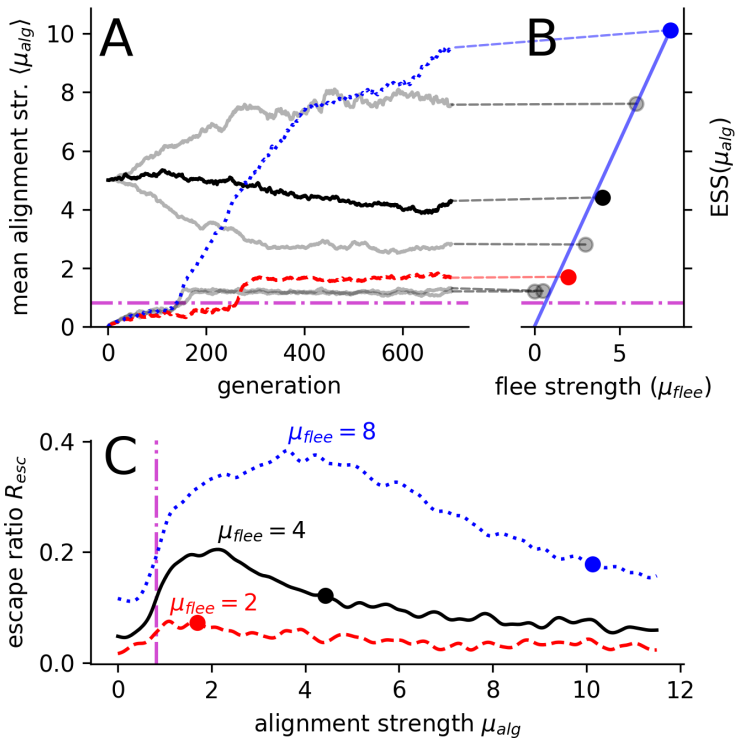

Figure 6.6: Evolution for different flee strengths μ_{flee}. **A**: evolution trajectory of the mean alignment strength μ_{alg} over 700 generations. **B**: dependence of evolutionary stable strategies (ESS) on μ_{flee}. Solid diagonal line shows the theoretically predicted linear dependence of the ESS on μ_{flee} assuming balancing of social and private information (see Sect. A.5). Dashed lines (**A**, **B**) connect the individual evolutionary runs (**A**) to the corresponding ESSs (**B**) obtained from multiple, longer evolutionary simulations. **C**: evolutionary stable states (circles) with respect to the group response, measured via the escape ratio R_{esc}, for three selected flee-strengths indicated with dashed, solid and dotted lines for $\mu_{flee} = [2, 4, 8]$ respectively (also color coded in all panels with red, black and blue). In all panels: the dash-dotted magenta line marks the order-disorder transition.

There is a clear dependence of the ESS on the flee-strength (Fig. 6.6A), more specifically the ESS exhibits a linear dependence on the flee-strength for $\mu_{flee} \geq 2$ (diagonal line in Fig. 6.6B). The order transition acts as a lower bound since the non-fleeing agents ($\mu_{flee} = 0$) equilibrate closely above it. Thus, the ESS for non-responding agents matches the group-level optimum due to the dynamical school structure at criticality.

The linear dependence on the flee-strength may be explained by prey balancing social vs. personal predator information. Social information about the predator is beneficial if the prey is in the second neighbor shell of the predator, i.e. where its neighbors but not itself respond directly to the predator. Thus, by coordinating with its informed neighbors it gains distance to the predator. However, if a prey directly senses the predator, social information of uninformed neighbors conflicts with its private information and therefore may hinder optimal evasion.

Therefore, individual prey agents should continue to evolve towards stronger alignment strength until the cost of the social inhibition of evasion counterbalance the benefit of social information. I find support for this conjecture by reproducing the observed linear dependence through a local mean-field approximation (see Sect. A.5) assuming the above balancing mechanism (Fig. 6.6B).

Coincides the responsive group optimum with the ESS?
Interestingly, also the escape ratio R_{esc}, as a measure of group response while controlling against spatial effects, exhibits a maximum in the strongly ordered region away from criticality (Fig. 5.13). This leads to the question whether the ESS coincide with the strongest group response. Indeed, the maximum of the escape ratio shows the same trend as the ESS of moving towards higher alignment strengths with increasing flee strength (Fig. 6.6C). But these maxima stay clearly below the corresponding ESSs (circles in Fig. 6.6C). This suggests that the system does evolve towards unresponsiveness (Torney et al., 2015) by increasing the social responsiveness above the optimum.

6.3 Robustness to parameter and implementation details

To ensure that the results are robust the evolution is repeated with (i) modified prey properties, i.e. changing the angular diffusion coefficient and introducing variable speed, (ii) a changed predator behavior, i.e. its agility, and (iii) changes in the evolutionary selection mechanism, e.g. by a predator which captures the prey during the simulation. The detailed robustness analysis is shown in the Appendix (Sect. 6.3). The general findings, that the phase transition is an evolutionary accelerator (i.e. it is unstable) and that the ESS shifts to larger alignment strength with larger flee strength is robust for all considered parameter and mechanism variations. I also checked another estimation of the ESS via the invasion fitness landscape that again reproduced the findings; however, it is harder to estimate the ESS from this landscape. An explicit variation of the implementation is the introduction of a heterogeneous environment. Since I intend with this variation of the selection algorithm to force the ESS to the phase transition, it is a rather unusual robustness approach. Therefore it is discussed explicitly in the next section.

6.4 An attempt to tune the ESS to criticality: evolution in a heterogeneous environment

In this section, I try to tune the ESS to the order-disorder transition in order to understand the necessary conditions for it. The observed ESS have always been in the ordered state also for all robustness test. That strongly suggests that the selection pressure of a moving predator is in the ordered parameter regime. In order to tune the ESS, another selection mechanism needs to act which favors the disordered phase. This additional selection mechanism introduces a heterogeneous environment, which is assumed to be a general important condition for the evolution to criticality (Hidalgo et al., 2014).

The self-sorting in this model predicts that high mortality of front individuals leads to an ESS in the disordered state. This mortality is implemented by declaring the most frontal prey dead. The extra selection is equivalent with the observed high risk of being in the front in the presence of sit-and-wait predators (Bumann et al., 1997). Since the current transition is close to the lower boundary of the alignment parameter ($\min(\mu_{alg}) = 0$), the transition is set to larger values, i.e. at $\mu_{alg,c} \approx 1.6$, by increasing the angular diffusion to $D = 1$ (ensuring that fluctuations allow equilibration in the disordered regime). Note that other environmental interactions, e.g. exploration and exploitation of food-sources, might simultaneously influence the fitness. Therefore, the front selection is only motivated by sit-and-wait predators but represents a selection mechanism that favors the disordered state.

The ESS decreases with increasing weight on the frontal-risk selection (Fig. 6.7). This seems plausible, however; in a similar study individuals evolved to criticality, if exposed to a diverse environment (Hidalgo et al., 2014). In fact the transition acts here as a fitness valley, marked by a zero-crossing of the fitness gradient with positive slope. It leads to multiple local optima (inset in Fig. 6.7), which only vanish if one of the selection mechanisms dominates. In consequence, for intermediate frontal-risk selection and predator selection exists two locally evolutionary stable states. (i) If the initial mean alignment is sufficiently below the transition in the disordered state, the ESS is at $ESS(\mu_{alg}) = 0$. (ii) Otherwise the ESS is in the ordered phase.

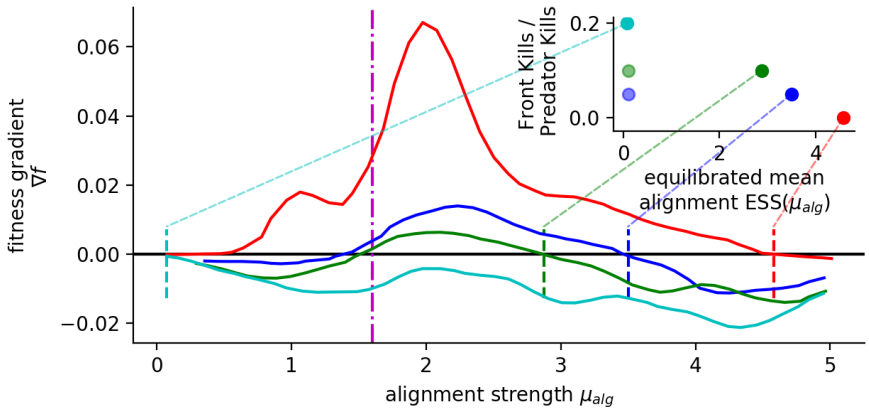

Figure 6.7: Evolution in a heterogeneous environment. Fitness gradients for different relative strength of the frontal-risk selection with respect to the simultaneously active predator-selection. In the frontal-risk selection the most frontal individuals are declared dead. The relative strength of the frontal-risk selection is defined by the ratio between agents killed at the front and by the predator, i.e. (Front Kills)/(Pred. Kills) $\in [0,\ 0.05,\ 0.1,\ 0.2]$. The evolutionary stable state (ESS) is defined by the zero-crossing of the fitness gradient with negative slope marked by a vertical dashed line. However, the lower bound is an additional ESS if the fitness gradient stays negative close to it which is marked by shaded points in the inset. Parameters are identical to the former simulations apart from the angular diffusion coefficient which is increased to $D = 1$ increasing the order-transition to $\mu_{alg,c} \approx 1.6$ marked by vertical dash-dotted magenta line. The flee strength is $\mu_{flee} = 4$.

In summary, to tune the ESS to the phase transition is non-trivial, that means I did not succeed with the "simplest" approach. It illustrates furthermore the robustness of the findings.

6.5 Summary

If an evolutionary algorithm allows the individual to reproduce according to its ability to avoid the predator, the system does not evolve to criticality. There are two main reasons: (i) not only the ability to directly avoid the predator is important but also the relative position in the collective and how likely the predator will visit this region and (ii) if a neighbor has information about the predator the best response is to respond as strong as possible to this information. The first reason causes the order disorder transition to be evolutionary unstable, even to be an accelerator of evolution. The second reason shifts the ESS far into the ordered state to the regime of socially highly responsive but signal unresponsive agents. The results are very robust to every variation I have tried. Even the attempt to "artificially" tune the ESS to criticality did fail.

7 Summary and discussion

I have shown, using a spatially-explicit agent-based model of predator-prey dynamics, that the group optimum with respect to predator avoidance is located in the vicinity of the critical point, in line with the criticality hypothesis. However, this optimality is not due to optimal transfer of social information, but rather to the highly dynamic structure of the group at the transition. Additionally, it shifts away from the phase transition for strong signals (large flee strengths). Yet, this group optimum at criticality for weak perturbations does not represent an evolutionary stable state under individual-level selection.

The work demonstrates the crucial importance of taking into account the self-organized spatial dynamics of animal groups when evaluating potential evolutionary benefits of grouping. It turns out that the mechanism responsible for the optimal collective performance (minimal capture rate) at the critical point, the highly dynamic and flexible structure of the collective, leads also to the steepest selection gradients in evolutionary dynamics, making the critical point evolutionary unstable. Evolution with random mutations enforces heterogeneity. In combination with the spatial symmetry breaking at the transition, it results in maximal assortative mixing and self-sorting close to the transition. These effects of self-organized collective behavior play a decisive role for the evolutionary dynamics close to criticality and "drive" the ESS out of the transition region towards the aligned state. In the presented model the ESS is in the strongly ordered phase, which suggests the evolution towards external unresponsiveness by overrating social information. Finally, we show that the ESS depends linearly on the flee strength, i.e. local perturbation strength. It can be explained by individuals that balance the benefit of social information about the predator's approach with the costs of useless social interactions if the information is directly available.

The transition on the collective level
In contrast to Hidalgo et al. (2014), the critical state in the presented model is not evolutionary stable, despite the similar setup: evolving agents that respond to conspecifics and to a changing environment (here the appearance of a predator). This can be explained by crucial differences to this work. Most importantly, in Hidalgo et al. (2014) each agent in isolation can already evolve to its "individual" transition by tuning its own gene regulatory network. This appears to be essential for a critical point corresponding also to the evolutionary stable state in their information-based fitness framework. In the here presented model, the disorder-order transition is a pure collective effect, i.e. individual agents can not exhibit any transition behavior by themselves. Furthermore, at the disorder-order transition, small differences in behavioral parameters translate into systematic differences in the self-organized spatial positioning within the group, which in turn directly influences the predation threat. This self-sorting (Couzin et al., 2002; Hemelrijk and Kunz, 2005; Jamie Wood, 2010) is maximal just above the transition and includes assortative mixing due to emergence of spatial "subgroups" with strong correlations between behavioral phenotype, spatial location and local school structure, which is potentially of interest in the broader context of collective task distribution and computation in spatially-explicit animal groups.

Feedback between structure and dynamics
There is another consequence of the tight coupling between local school structure and individual dynamics: The extent of the collective is largest at the transition because the responsiveness to directional fluctuations is maximal, i.e. local fluctuations induce deviations in the movement of different parts of the school causing the school effectively to expand. In systems with an one-way influence from structure to dynamics (fixed networks), it is known that at the order-transition structural differences cause the largest dynamic variability (Nykter et al., 2008). We show that in a system with additional feedback from the dynamics to the structure, also the structure has the highest variability at the transition, which may have important consequences for collective

computations (e.g. enhancing the collective gradient sensing Berdahl et al., 2013; Hein et al., 2015). It shows that interactions on fixed (Brush et al., 2016; Chicoli and Paley, 2016; Vanni et al., 2011) or randomly rewiring (Torney et al., 2015) lattices might miss these functionally highly relevant features of collective behavior.

Model alternatives
For simplicity I used here an established social force model (Couzin et al., 2002; Harpaz et al., 2017; Hemelrijk and Kunz, 2005), which also received empirical support (Calovi et al., 2018; Katz et al., 2011). As highlighted in Sect. 3.3, the analyzed model is one of the simplest models with topological interactions and volume exclusion. Alternative social interaction mechanism may also lead to highly ordered collective movement (Bastien and Romanczuk, 2020; Jhawar et al., 2020; Lei et al., 2020; Romanczuk et al., 2009), but the crucial spatial effects such as self-sorting and assortative mixing are generic and will be independent on the details of social interactions. Still, it is possible that variations in the alignment interaction alter the positions of the ESS (but not with respect to the critical). For example, an interesting extension for the alignment is to be sensitive to the curvature of the trajectory (Cavagna et al., 2014). It reproduces the linear propagation of directional information in starling flocks.

Another key model mechanism is the flee behavior that is modeled as a repulsion away from the predator. An agent that is not able to change its speed has only the option to get away from the trajectory of the predator. However, if a simple repulsion force is assumed that directly points away from the predator, only the prey-agents at the side are able to escape the predator. The agents directly in front stay in front because they escape to the front. That the most endangered individuals, the individuals in the predator's front, behave worst by following the flee-force can be argued to be unrealistic. In fact, experiments suggest that individuals do not flee directly away, but rather follow a certain angle (Domenici and Batty, 1997). I have tested this possibility (see Sect. A.8) and verified that the direct repulsion is worse than a different flee angle. However, as long as the flee response increases the survival probability, social information is beneficial and therefore should the observed results be robust against an improved fleeing mechanism.

Group response dependence on signal strength
Another key finding of this thesis is that the group responsiveness depends strongly on the signal strength. The word responsiveness is used in the attempt to explain the biological relevance of physics motivated measures as susceptibility (Ballerini et al., 2008; Mateo et al., 2017; Vanni et al., 2011). I find that biological relevant group measures of response shift their maximum away from criticality for larger flee strength. This happens as well for the structurally non-controlled measures (Sect. 5.4.1). That means, that for strong signals the group responds best in the ordered region and not at the order-disorder transition. This is a vital point, because it stresses that the responsiveness of a system estimated by its internal fluctuations is only fair if the signal is weak, but loses validity for strong signals. A possible reason is that strong signals perturb the system far away from their initial steady state. A local adaptation to the signal could be accelerated if the system is far from the transition point. There, dynamical systems have a larger Lyapunov exponent consequently a faster relaxation. Thus, the susceptibility should be taken with caution in the analyzes of response.

Group and individual optimum in the ordered state
As stressed above, the group optimum shifts to the ordered state for large flee strength. However, also at low flee strength the response measures that control against spatial structure show an optimum in the ordered phase. This finding agrees well with the study by Wood and Ackland (2007), who found that the predator success is reduced in the polarized school. Interestingly, they did also identify the ESS in the highly ordered region, in agreement with the findings in this study. Torney et al. (2015) also found that evolution results in an ESS with a strong weight for social information, much stronger than the social weight which corresponds to the group optimum. They used a decision making model of cooperating agents with a random (also in time) or fixed interaction network. Each agent can access information about the environment by a private estimate and/or

rely on the information of its neighbors. The analogous findings of this thesis and both studies (Torney et al., 2015; Wood and Ackland, 2007) suggest that it is a common feature of system consisting of cooperating units to evolve to a state where social information is overrated. As a consequence the system is less responsive to environmental information.

In this study the flee strength was identified as key parameter for both, the group optimum and the evolutionary stable state. The question emerges: which flee strength is realistic? In this model an agent considers still all social-forces if it encounters the predator. However, why should a biological agent behave in this way? A more realistic assumption is that prey ignores all social cues when it detects the predator. This would correspond in the here presented model to an infinite flee strength. However, volume exclusion should still hold, e.g. a repulsive force should exist. If such a system would evolve in parameters, the ESS would shift to the extremely ordered region. There would be no deficit in paying attention to social cues anymore (alignment cues can not corrupt the infinitely strong flee force). Moreover, that is maybe how startling cascades in fish evolved (Rosenthal et al., 2015; Sosna et al., 2019).

Alternative fitness influencing mechanism/contexts
The finding that, individual-level adaptation does not evolve the prey groups towards criticality, suggests that evolutionary adaptation alone is not a sufficient mechanism for self-organization towards criticality. Whereas it does not exclude the possibility that animal collectives may operate in the vicinity of phase transitions in order to optimize collective computations, it clearly demonstrates the necessity for further research on biologically plausible proximate mechanisms of self-organized criticality in animal groups. A general, fundamental difficulty is that besides predator evasion there are various ecological contexts and other dimensions of (collective) behavior that will affect individual fitness. Here, by focusing on a dominant selection pressure, namely predation, I neglect other mechanisms, as for example resource exploration and exploitation (Brush et al., 2016; Hein et al., 2015; Monk et al., 2018; Wood and Ackland, 2007) whose ESS can also depend on the resource abundance (Brush et al., 2016; Monk et al., 2018; Wood and Ackland, 2007). This emphasizes the importance to study collective behavior in the wild (Francisco et al., 2020; Graving et al., 2019; Handegard et al., 2012; Hansen et al., 2020) to provide more empirical input on actual relevant behavioral mechanisms as well as variability of behavior across different contexts. However, I have shown that even by combining two opposing selection mechanisms (see Sect. 6.4), which on their own favor ordered or disordered state respectively, the critical point does not correspond to an evolutionary attractor, it remains an evolutionary highly unstable point. Since the group optimum for weak signals is at the phase transition, a multi-level selection mechanism (within- and between-group selection, Wilson, 1975, 1997) could drive the collective to criticality. However, also with multi-level selection the phase transition is only a possible ESS if the flee strength, the signal, is weak. Still, it would drive the system closer to criticality compared to the examined pure within-group selection mechanism.

Other phase transitions
I focused here on the prominent directional symmetry breaking transition. However, recently it was suggest that a transition in the speed relaxation coefficient may represent a functionally relevant critical point in flocking behavior (Bialek et al., 2014). Individuals with lower relaxation constants are less bound to their preferred speed and may gain fitness benefits due to their ability to adapt faster to higher speeds of fleeing conspecifics. Consistent with this hypothesis, guppies (*Poecilia reticulata*) exhibit stronger accelerations in high-predation habitats (Herbert-Read et al., 2017). However, fish exhibit a reflex-driven escape response, so-called startle, which was recently shown to spreads through fish schools as a behavioral contagion process (Rosenthal et al., 2015; Sosna et al., 2019). This suggests that at least in the context of collective predator evasion in fish, another type of a critical point may be highly relevant, which is analogous to the critical threshold in epidemic models. It separates states of non-propagating startle response, with only small localized response of few individuals, from avalanche-like dynamics, where a single fish may cause a global startle cascade. Note that startle response is a vital component in contexts where fish or other animals are

moving at an average speed that is lower than their maximal swimming speed. However, if animals are already escaping, they are likely to be close to their maximum speed and can physically not startle anymore (Hansen et al., 2020, e.g. sailfish or marlin attacking sardine schools,). Thus, the fixed speed scenario analyzed in this thesis is a fair approximation for the latter context.

For the behavioral contagion process the network is derived from the visual interaction network (Rosenthal et al., 2015; Sosna et al., 2019). There, the local to global contagion transition can be reached by changing the strength of the links in the network. However, another percolation like transition that also enhances the group response is tuned via the average degree of the network (Mateo et al., 2017). Mateo et al. (2017) find that agents in a Vicsek model avoid a predator best if they interact with about 20 of their closest neighbors. The question arises how relevant this transition is for real flocks, i.e. have individuals the cognitive ability to process the information of 20 neighbors simultaneously and are 20 neighbors accessible from the visual field? Importantly, Mateo et al. (2017) use no volume exclusion, which should raise concern about the relevance of their findings.

Hein et al. (2015) found that a collective on the search for resources evolves to a first order phase transition. This is interesting because first order transitions are not associated with a diverging susceptibility. Thus, they do not classify as possible candidates of the criticality hypothesis. Individuals which are close to the transition where able to switch between socially unresponsive exploratory behavior (high speed if no resource is detected) and collective cohesive behavior. They exploited the resource (speed is lowered if resource is detected) depending on the local resource abundance (Hein et al., 2015). This gave them an advantage against prey who could only stay in either state. It is an interesting example for how a phase transition can be beneficial without relying on the susceptibility as an argument.

Final remark

Overall, our study does not reject the general possibility that animal groups manifest critical behavior and that it may be adaptive. However, it highlights importance of identification of biologically plausible proximate mechanisms for self-organization towards - and maintenance of - critical dynamics in animal groups, which account for spatial self-organization and the corresponding ecological niche.

Part II

Vigilance as driver for cohesion differences between zebrafish selection lines

In contrast to the previous chapter, which was purely theoretical, this chapter combines analysis of experimental data and modeling of shoaling behavior of different experimentally created selection lines of zebrafish (*Danio renio*). The project was initiated by the Cooperation and Collective Cognition Network (CoCCoN), an interdisciplinary network between Princeton University and Humboldt-Universität zu Berlin. The framework facilitated the contact between different disciplines and was the driver for this collaborative project with experimental biologists and their study system. A major fraction of this part of the thesis is available as preprint (Sbragaglia et al., 2020).

To prevent confusion please be aware of the following terminology. The selection line in which the large individuals were preferentially removed in consecutive generations is referred to as **large-harvested** line (**LH**). To the selection line with no size preference for removal, I refer to as the **random-harvested** line (**RH**). The remaining selection line, in which small individuals were preferentially removed in consecutive generations, is referred to as **small-harvested** line (**SH**).

In this part of the thesis, I first give a brief introduction to the possible effects of fishing/fisheries on collective behavior and the relationship between risk and cohesion and vigilance and how it changes in the group. Second, I introduce the size-harvested zebrafish selection lines and how they differ in cohesion, risk taking and other measures. Third, I introduce a burst and coast model that is motivated by a possible explanation of unexpected behavioral differences between selection lines. Forth, the proposed model is fit to each selection line to investigate if the possible explanation is supported by the data. Fifth, the model representation of the selection lines are used to generate predictions based on simulating the selection lines in different contexts (natural predation and two fishing scenarios). Finally, a discussion and summary of this part is provided.

8 Introduction: fishing, vigilance, risk and cohesion

Recreational fishing and industrial fishery can have a tremendous impact on the mortality rate of fish. In a natural context small size juveniles have a higher mortality rate (Lorenzen, 2000; Sogard, 1997) but exactly this changes under strong fishing pressure (Heino et al., 2015). A clear consequence is the selection for fast life histories, i.e. earlier maturation (Heino et al., 2015). Fishing can also have an impact on the onset of migration, if it is not distributed evenly over the year (as shown in sockeye salmons; Quinn et al., 2007).

In principle, fishing has been simulated in experiments by removing large individuals (Heino et al., 2015). This is a reasonable assumption for recreational fishery and commercial fishing because removing large individuals is typical in most fisheries (Salomon and Holm-Müller, 2013). However, there are other possible reasons that an individual's size influences its mortality. For example, an invasive predator could selectively prey upon smaller individuals (e.g. lionfish, *Pterois miles*, in the Mediterranean Sea; D'Agostino et al., 2020). Or diseases released from temperature constraints due to climate change Karvonen et al. (2010) could affect specific size classes. In this context, it has been predicted that fish populations increase in shyness if the mortality of large individuals is strongly increased (Andersen et al., 2018).

It is in general interesting how individual behavior influences the shoaling behavior. If the above mentioned predictions are true, i.e. that an increased mortality of large individuals increases the shyness, a reasonable consequence for the group level behavior is an increase of cohesion to benefit from collective predator protection (Krause and Ruxton, 2002). This hypothesis would be in line with previous studies: Fish in high predation habitats are more cohesive compared to fish of the same species from low predation habitats (Herbert-Read et al., 2017), shoals respond to an alarm substance with an increased cohesion (e.g. Sosna et al., 2019; Speedie and Gerlai, 2008) and reduce by cohesive behavior their zone of danger (Hamilton, 1971). However, only because observation and theory agree that an increased cohesion increases the survival probability in risky environment, it does not necessarily mean that another trait, e.g. shyness, which also increases the survival probability, causes the cohesion to increase.

Shyness can be directly associated with vigilance e.g. more shy kangaroos (those with a larger flight initiation distance) are also more vigilant (head is up more frequently) (Edwards et al., 2013). However, it is largely unexplored how the individual vigilance affects the group behavior. On the other hand, some studies suggest that group size affects vigilance, leading to an increased vigilance in larger groups (Godin et al., 1988; Ward et al., 2011). This is however not necessarily always the case, as for example territorial aggression between conspecifics reduces vigilance (Hess et al., 2016). While the effect of group size on individual vigilance is well studied in species in which the level of individual vigilance can be easily approximated, e.g. in general species that lift their head if they are vigilant (Beauchamp, 2003, 2008; Beauchamp and Ruxton, 2008; Edwards et al., 2013), it is very difficult to estimate the level of individual vigilance in fish. Therefore, Ward et al. (2011) could only hypothesize that a fish in a group scans a smaller portion of the environment more frequent, i.e. they could not measure it.

Here, I will check if, in principle, vigilance could even result in a decrease in cohesion. More vigilant individuals scan the environment more frequently but might miss social information and therefore be less cohesive. This idea builds up on an assumed attention trade-off between social and environmental information. For example, cooperatively breeding cichlid fish respond with a longer delay to a predator if they engage in territorial aggression (Hess et al., 2016). In three-spined sticklebacks, *Gasterosteus aculeatus*, the more aligned individuals responded slower to an external cue (MacGregor et al., 2020), i.e. the individuals who pay more attention to social cues are slower in detecting external cues. Similar results exist for herring, *Clupea harengus* where solitary

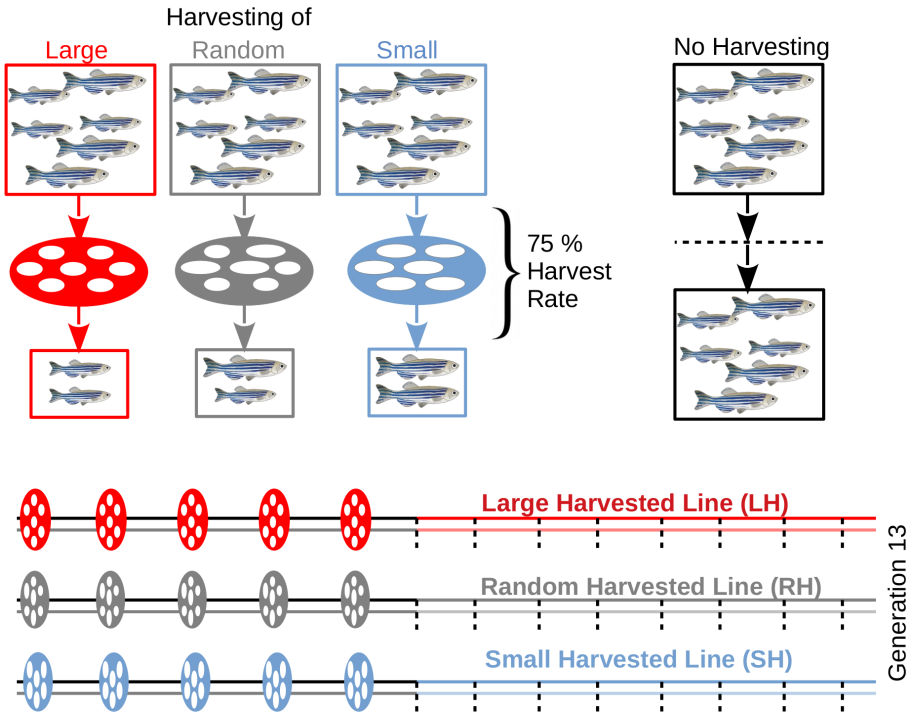

Figure 8.1: Size-selective harvesting was repeated for the first five generations. Each harvesting event is illustrated by a filter with colors red, gray and blue representing large-, random- and small-harvesting events, respectively. At each event the harvest rate is 75 percent. No harvesting was applied after the 5th generation for another 8 generations. The data was collected at generation 13. There exist two samples (replications) of each selection line illustrated via the transparent lines below the solid lines.

individuals responded faster to an external cue than schooling individuals (Domenici and Batty, 1997). Therefore, the attention trade-off is not only plausible (limited perception and cognitive capacities) but also supported by experiments.

In more general terms, the main research questions of this second part of the thesis are

- Which behavioral mechanism is altered by strong fishing pressure?
- Can these changes in behavior influence the survival against natural predators and fishing scenarios?

While I have emphasized that behavioral changes due to fishing pressure are possible, the second question focuses on the consequence of these behavioral changes in different contexts.

9 Large-, random- and small-harvested zebrafish selection lines

The selection lines were under size-selective harvesting pressure in the first five generations with a harvesting rate of 75 percent (Fig. 8.1). Thereby, resulting in two selection lines for large harvested fish(**LH**), random harvested fish (**RH**) and small harvested fish (**SH**). The two replications per selection line are meant to rule out effects caused by genetic drift. Thus, if only one replication shows an effect, it could be unrelated to the selection process and caused by the small population size and random selection for traits.

Each selection line at generation 13 was randomly split in 6 groups of 8 fish, i.e. 36 groups in total. The behavior of these groups was tested in two different contexts. In a deep tank, the risk-taking behavior was assessed by the time the fish spent at the surface. There the tank was recorded from the side. In a shallow tank, the shoaling behavior was estimated with tracked video recordings from above.

This was a collaborative project with Valerio Sbragaglia as the main experimental collaborator. Before I analyze the data I clarify my contributions. After that, this chapter is split in four parts. First, I present how the experimental data was processed in order to compute group and individual averages. Note that only the data of the shoaling experiment was processed in this way, i.e. data of the risk-taking experiment are from a previous study (Sbragaglia et al., 2019). Second, I will present results of the risk-taking experiment. It served as motivation and was the initiator of the collaborative project in the framework of CoCCoN and the working-hypothesis of this chapter. Third, the group averages and the individual averages of the shoaling experiment are analyzed. Forth and at last, I discuss the results and present a possible explanation for why more risk-taking groups are also more cohesive.

9.1 Contributions

This work was collaborative (Sbragaglia et al., 2020). Valerio Sbragalia was responsible for designing and performing the experiments, video recording, tracking and statistical analyses. My main contributions were processing and analyzing the tracking data, agent-based model development and implementation and systematic model simulations. I was provided with the tracking data from id-Tracker (Pérez-Escudero et al., 2014), whose output was further processed by me. The interpretation of individual and group measures of the selection line in Sect. 9 and the formulation of the working hypothesis was done together with all authors of Sbragaglia et al. (2020). None of the statistical significance tests in Sect. 9 were done by me.

The data from the risk-taking experiments presented in Sect. 9.3 are taken from (Sbragaglia et al., 2019). In the rest of this part of my thesis, I was the main contributor; however, the conceptual design of the simulations to predict the adaptive value of the different selection lines in Sec. 12 was done collaboratively by all authors of Sbragaglia et al. (2020).

9.2 Data processing

The software id-Tracker (Pérez-Escudero et al., 2014) provides trajectories of each individual without the need tag them. It identifies individuals based on consistent differences in coloration or size (relative pixel values) that might be hard/impossible to distinguish by eye. Next to the position data, also the probability of the correct identification p_{ID} is provided. For most group measures it is irrelevant that the individual is labeled with its correct ID. For example, the mean nearest neighbor distance $\langle \mathrm{NND} \rangle = \sum_{ID} \mathrm{NND}_{ID}$ is independent of correct identification. However, for

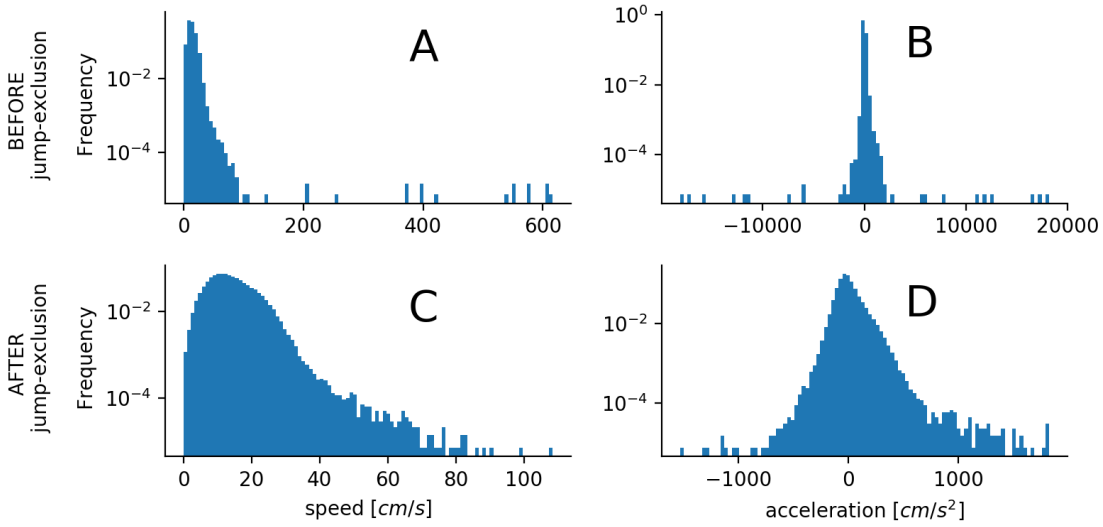

Figure 9.1: Histogram of speed (**A**) and acceleration (**B**) of individual fish in a group 8 recorded for 5 minutes. Only data is excluded that had a low identification probability (no jump-exclusion). After also jumps, strong acceleration directly followed by deceleration of similar magnitude, are excluded the distribution of speed (**C**) and acceleration (**D**) have less extreme values and a more defined tail.

every measure based on the change of positions or for the estimation of individual behavior, the correct identification is pivotal.

Therefore, for the group-measures that are unrelated to individuals or position changes, the data is filtered by excluding frames with at least one negative identification probabilities ($p_{ID} < 0$, i.e. all individuals must be detected but their correct ID is irrelevant). For a selection line average of a group-measure, each group average contributes equally to the line average. The contribution of each group is not weighted by its fraction of valid frames because the tracking quality must not influence the group-measures. In order to avoid that groups with low number of valid frames (weak estimates) influence the line average, only the 10 groups with largest fraction of valid frames are included in the analysis for each line (i.e. 10 groups per selection line and 5 groups per replica).

For the individual measures or measures based on positional changes, all trajectories are excluded that have identification probability below 85 percent ($p_{ID} < 0.85$). To further exclude identity switches or mal-tracked data, all unrealistic jumps are excluded. The jump-exclusion is intended to exclude obvious tracking errors or missing frames while preserving fast acceleration startle responses. An allocation caused by tracking errors has two specific characteristics: (i) the fish is located in a different position than the former speed would have predicted, and (ii) the speeds before and after the dislocation are similar. In other words, the fish accelerates strongly with a_0 and directly at the next frame decelerates with a_1 back to about its previous speed. Thus, a part of the trajectory is declared as jump if all three conditions below are fulfilled:

- an acceleration is followed by a deceleration: $0 < a_0,\ a_1 < 0$
- acceleration and deceleration are similar: $-0.8a_1 < a_0 < -1.2a_1$
- the acceleration is strong, which is the case if $a_0/v_{-1} > 2$, i.e. the speed triples at least.

These parts are declared as jumps and excluded from the analysis. However, a zig-zag track is not detected by this criteria. Zig-zags tracks are characteristic for zebrafish, but not on the timescale of frames (i.e. milliseconds). A part of a trajectory is declared as zig-zag and excluded if:

- the heading change at two consecutive frames is large: $|\Delta\phi_0| > \pi/2,\ |\Delta\phi_1| > \pi/2$

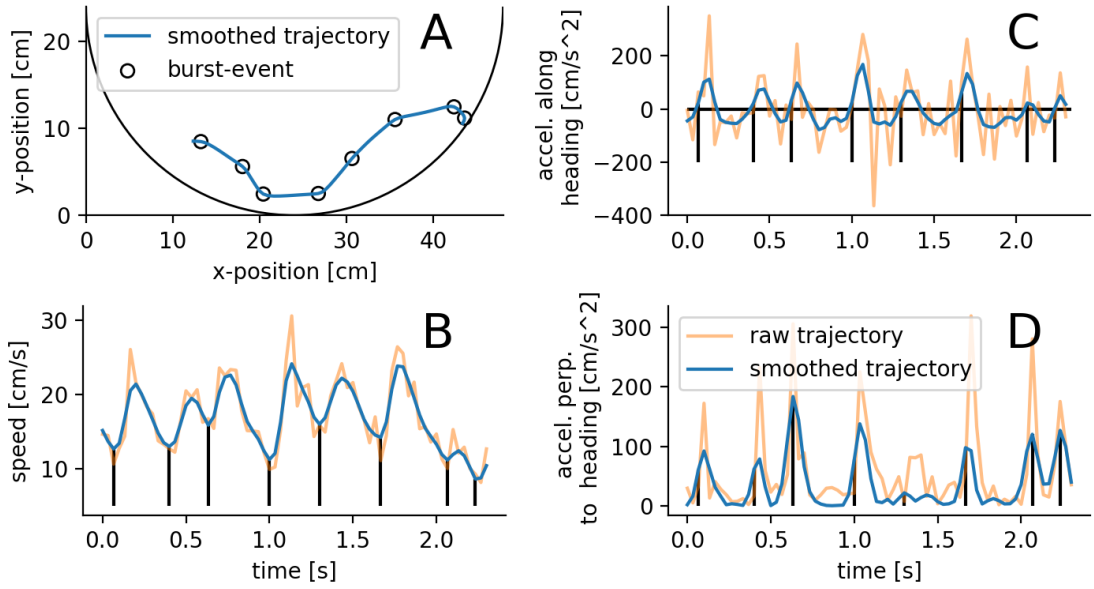

Figure 9.2: Example smoothed trajectory of a swimming zebrafish generated from ID-Tracker data (**A**). The large circle represents the tank-boundary. For this trajectory the speed (**B**), acceleration along (**C**) and perpendicular (**D**) to heading direction are shown computed from the raw (orange) and smoothed (blue) trajectory. The black circles (**A**) and vertical black lines (**B, C, D**) indicate burst events, i.e. the onsets of a positive acceleration period of the smoothed data (**C**). For the smoothening a moving window average with Gaussian kernel was used. The kernel width $\sigma_{smoo} = 1.13$frames was estimated by comparing manual counts of burst events with automatically detected (see Fig. B.2).

- the changes in heading happen at least at the mean-movement speed: $v_0 > \langle v \rangle,\ v_1 > \langle v \rangle$.

The second conditions assures that parts with normal tracking accuracy are not excluded, zig-zag is allowed at low speeds (it is smoothed out later).

This jump- and zig-zag-exclusion removes outliers from the speed and acceleration distribution (compare Fig. 9.1A, B with C, D) and allows arbitrarily large accelerations (Fig. 9.1D). To ensure that the remaining large accelerations and speeds do not correspond to tracking errors, I checked the trajectories of the thirteen largest accelerations (Sect. B.1.1). Only one of the thirteen is caused by a tracking error and was not excluded because the dislocation is small (small acceleration). Surely, a better exclusion criterion is possible; however, it is already a good compromise between keeping as much data as possible and excluding obvious tracking-errors.

To remove the remaining tracking imprecisions I applied a moving average on the position data with a Gaussian kernel of width $\sigma_{smoo} = 1.13$frames. The width was defined by comparing the number of detected burst-events, i.e. the number of periods of positive acceleration, with a manually counted number of bursts in a randomly selected trajectory of 2000 frames. The smoothed velocities conserve the burst-dynamics of the raw data (Fig. 9.2B). The smoothed acceleration along and perpendicular to the heading direction are clearly lower than the accelerations of the raw data (Fig. 9.2C, D). This is to be expected and likely caused by the tracking-noise. However, please note that the acceleration perpendicular to the heading direction, i.e. the turning, starts always before the acceleration along heading direction. This observation will be important in Sect. 11.1.2 for estimating the burst force.

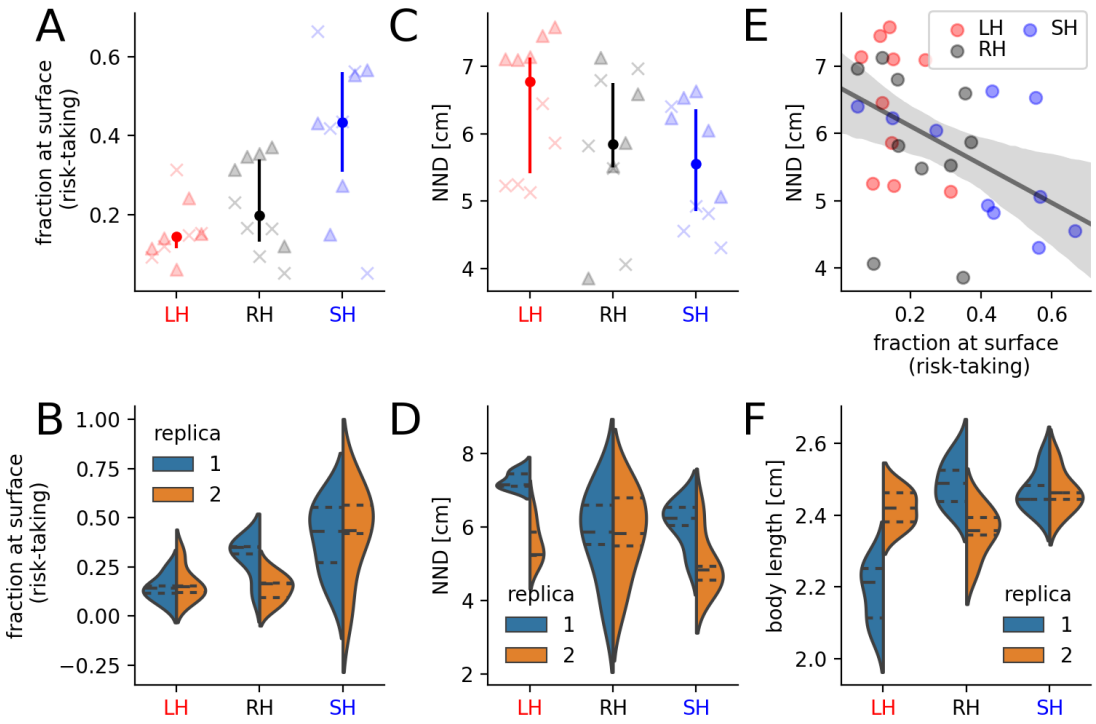

Figure 9.3: Group averages are shown for the risk-taking, estimated via the average fraction of time spent at surface (**A, B**), and the nearest neighbor distance (NND, **C, D**). The correlation between these two measures is negative, independent of the selection line (**E**). Differences in body length between the selection lines show a clear replica dependence for the LH line (**F**). **A, C**: triangles and crosses represent group averages of replica 1 and 2 respectively, dots and vertical lines represent quartiles of all groups combined. **A, C, E**: colors red, black and blue represent the large-, random- and small-harvested lines, respectively. **B, D, F**: splitted violinplots with each half representing the histogram over the group-averages corresponding to one replica line.

9.3 Risk-taking behavior

Between 230-240 days post fertilization, the risk-taking behavior was tested by measuring the time the fish spent at the surface (Sbragaglia et al., 2019). Each group of fish was moved into an experimental tank (width x length x height = (10 x 30 x 25) cm^3) with water height of 22 cm. The groups were left undisturbed for only 3 minutes, which is assumed to be sufficiently long to avoid having the typical diving responses (Sbragaglia et al., 2019). However, 3 minutes are clearly not enough for acclimatization (Kleinhappel et al., 2019), and therefore the groups are most likely stressed in which case they tend to avoid risky environments as the surface. After 3 minutes a small amount of food was added at the surface and the behavior was recorded for 30 seconds. The fraction of time spend at the surface was measured, which is the average fraction of time a fish spent in the upper third water column (first 7cm below the water surface). The LH line spends a smaller fraction of time at the surface than the RH line, and the contrary for SH line (Fig. 9.3A). Note that this is consistent in between the replicas of LH and SH selection line, only the RH line shows a deviation between replicas (Fig. 9.3B).

9.4 Shoaling behavior

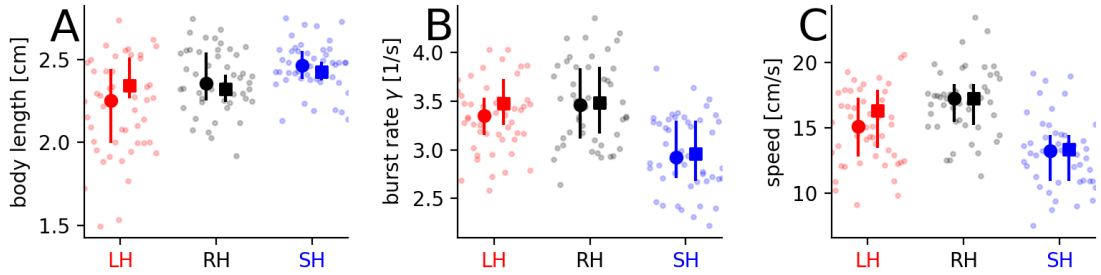

Figure 9.4: Selection lines individual level averages of the body length (**A**), burst rate γ (**B**) and speed (**C**). All panels: each small transparent circle represent the average value of one individual. The large circle with vertical lines shows the median, and quartiles of considering all individuals. The large rectangle with vertical lines shows the median, and quartiles of considering a size-matched subset of each selection line.

9.4.1 Experimental setup

At 190 days post fertilization, we measured shoaling behavior in groups of 8 zebrafish in a white round arena (diameter of 49 cm) with 10 cm of water. The arena was placed on a table behind a white curtain to minimize disturbance to the fish during the experimental trials. We recorded behavior using a webcam (resolution: 1920 x 1080 pixels; frame rate: 30 frames per second) from about 1m above the arena. Zebrafish were introduced in the experimental arena and left undisturbed for 25 min before starting the experimental trial. Video recording lasted 5 min and after that we measured the standard length of each fish on a petri dish with millimeter paper anesthetizing the fish using a clove oil dilution in ethanol and water. The data was processed as described in Sect. 9.2.

9.4.2 Results

With respect to the polarization $\Phi = |\sum \vec{v}_i/v_i|/N$ the selection lines show no significant differences ($\Phi_{LH} = 0.40 \pm 0.06$, $\Phi_{RH} = 0.41 \pm 0.05$, $\Phi_{SH} = 0.39 \pm 0.07$). The mean nearest neighbor distance shows a clear trend if it is measured in centimeter. Thereby, the LH line is less and the SH line more cohesive than the RH line (Fig. 9.3C). This trend becomes significant if the metric body-length (BL) is used (Fig. 9.3 F). If the cohesion is estimated via the mean inter-individual distance (IID), the trend is significant for both length units (centimeter and BL). However, the IID is a less robust estimator for cohesion because it increases strongly if the group splits in two subgroups, even if each group is very cohesive. The interesting point about the stronger cohesion of the LH line is that it showed a decreased risk-taking behavior (compare Fig. 9.3A and C). This seems to be a general relation also present within the other selection lines, i.e. groups of the SH line are more risk-taking and also less cohesive (Fig. 9.3E). If the two replica lines are compared, the NND of the LH line is larger only for replica 1 (Fig. 9.3C, D). Therefore, it could be due to a genetic drift in the selection lines. However, the body length of replica 1 is decreased with respect to control (Fig. 9.3F), as expected for the large-harvested line. This suggests that replica 1 of the LH line is more reliable than replica 2 and therefore further supports the trend between selection lines of the NND.

The selection lines further differed in individual measures as in the already mentioned body length (Figs. 9.3F and 9.4A). To ensure that the differences in other individual measures are not caused by known size effects (Bainbridge, 1958), the statistical analysis was repeated for a size-matched subset of individuals from each selection line. The body-lengths of the size-matched subset do not significantly differ in between selection lines (rectangles in Fig. 9.4A). The differences in average burst-rate γ, i.e. the number of periods of positive acceleration per time, and in the strongly related average individual speed are the same for the whole dataset and the size-matched

subset (compare rectangles with circles in Fig. 9.4B, C). The RH line is for both measures larger than the LH line, which in turn is larger than the SH line, i.e. the lowest (fastest) speeds and burst rates are detected for the SH (RH) line.

9.5 Possible explanation: vigilance and attention trade-off

It seems that the more risk-taking groups have a stronger cohesion (Fig. 9.3A-E). However, cohesion is considered one of the major defense mechanisms against predators (Krause and Ruxton, 2002). This consideration is supported by theoretical arguments, as a decreased zone-of-danger for more cohesive schools (Hamilton, 1971), as well as by experimental findings, e.g. guppies living in habitats with increased predation pressure are more cohesive compared to guppies from low predation habitats (Herbert-Read et al., 2017). Therefore, it is expected that more risk-taking groups are less cohesive. This is the exact opposite of what the zebrafish selection lines show. This is in principle possible, because the selection lines were not under natural predation and consequently the behavioral traits could have evolved in an unnatural manner.

From a theoretical point of view I assume that there is always noise on the individual behavioral decisions. The motivation of this noise is manifold. For example, its cause can be the imperfect estimation of the neighbors velocity and speed or a movement motivation unrelated to social cues (e.g. possible food-source, individual preference of light intensity). In the context of predator avoidance, behavioral noise, i.e. the deviation from expected social behavior due to response to random environmental cues, could be associated with vigilance. Vigilance is normally quantified by how often the head is up during feeding (Beauchamp and Ruxton, 2008). However, vigilance in fish is estimated for example by the probability that an individual detects an alarm signal, i.e. if it startles (Godin and Morgan, 1985). If no alarm signal is present, more vigilant individuals are expected to react more often to random environmental cues, as the reflection on the water surface or a water splash. Thus, more vigilance equals more noise, if no cue is present. More noise, in turn, can lead to less cohesion.

In summary, a possible explanation for the unexpected cohesion differences in between selection lines is: The less risk-taking groups (LH line) are more vigilant which lowers their cohesion because they substitute social behavior by responses to random environmental cues.

10 The burst-coast model

The experimental data from the previous chapter revealed an unexpected trend in cohesion with respect to the risk-taking behavior of the groups. I presented a possible explanation, which hypothesizes that vigilance can effectively introduce noise because more vigilant individuals respond more often to random environmental cues (in the absence of a real alarm cue), compared to less vigilant conspecifics. To check if the hypothesis works mechanistically, I propose in the first section of this chapter an agent-based model in which agents either follow (random) environmental cues or social cues. In the second section, the presented model is compared to existing burst-and-coast models. The next chapter fits the model to each selection line, i.e. creates respective model representations, and finally checks if the proposed mechanistic explanation for the differences in cohesion is reasonable.

10.1 Model description

The section is structured in, first, describing the implementation of the burst and coasts, typical for zebrafish. Second, I introduce the social forces. Third, the environmental force is presented and the detection of environmental cues, i.e. how agents detect threatening cues, as a predator or fishing agent. Forth, the wall avoidance mechanism is described. The wall avoidance is necessary because the model is later compared/fitted to experiments in which the tank boundary limits the exploration area.

10.1.1 Burst and coasts

The burst-coast model intends to mimic the burst-coast swimming behavior of zebrafish (*Danio rerio*). I assume that a fish is accelerating only during the burst phase with a constant force of magnitude f_b, while no forces are present during the coast phase, implying a deceleration due to friction. The differential equations of motion defining current position $\vec{r}_i$ and velocity $\vec{v}_i$ of an agent i are:

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i \tag{10.1a}$$

$$\frac{d\vec{v}_i}{dt} = -\beta\vec{v}_i + \vec{F}_i(t) \tag{10.1b}$$

with $\vec{F}_i(t)$ as a finite social or environmental force vector with $|F(t)| > 0$, for fish in the bursting phase. A fish decelerates passively during the coasting phase, i.e. the force vector vanishes $|\vec{F}_i| = 0$. β is the friction coefficient. The velocity change of Eq. 10.1 can be split into the part parallel and perpendicular to the current velocity direction $\vec{v} = v[cos(\varphi), sin(\varphi)]$:

$$\frac{d\vec{v}_i}{dt} = \frac{dv_i}{dt}\vec{e}_{\parallel,i} + v_i\vec{e}_{\perp,i}\frac{d\varphi_i}{dt} \tag{10.2a}$$

$$\frac{dv_i}{dt} = -\beta v_i + \vec{F}_i \cdot \vec{e}_{\parallel,i} \tag{10.2b}$$

$$\frac{d\varphi_i}{dt} = \frac{1}{v}\vec{F}_i \cdot \vec{e}_{\perp,i}\ . \tag{10.2c}$$

Note that the v^{-1} dependence of the angular change (turning rate) was, next to being directly derived from Eq. 10.1, verified by the tracking data (Fig. B.3).

The burst behavior is defined by the burst rate γ, the burst time or duration of the burst t_b, and the burst force $\vec{F}$. In particular, the burst force $\vec{F}$ governs whether the fish uses social or

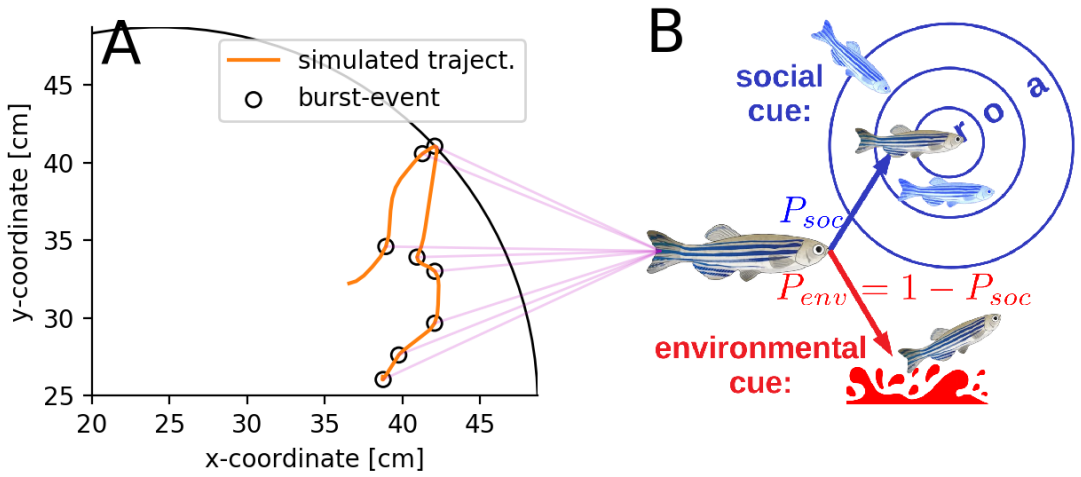

Figure 10.1: A: Simulated trajectory of 2 seconds with circles indicating the detected burst events. There might be more burst events, because a burst can directly follow another burst that is only detected as one long burst. **B**: At each burst event an agent probabilistically decides if it either follows social (with probability P_{soc}) or environmental cues (with probability $P_{env} = 1 - P_{soc}$). If the agent follows social cues, it updates its position according to its neighbor states and to which zone (r=repulsion, o=orientation, a=attraction) they belong to. In case of environmental cues, the agent's burst force points in a random direction. The randomness is motivated by random environmental perturbations, e.g. a reflection on the water-surface, unrelated to danger.

environmental cues. The fish decides at the start of a burst with a probability P_{social} if reacting socially to the other fish or, with probability $1 - P_{social}$, to environmental cues.

The resulting burst force $\vec{F}$ is

$$\vec{F} = \begin{cases} F\hat{f}_{soc}, & \text{if } p < P_{soc} \text{ with } p \sim U(0,1) \\ F\hat{f}_{env}, & \text{otherwise} \end{cases} \tag{10.3}$$

with F as the magnitude of the burst force and $\hat{f}_{env}$, $\hat{f}_{soc}$ as the unit vectors in direction of the environmental or social cue.

10.1.2 Social force

The social force is motivated by a three zone model (Couzin et al., 2002) consisting of a repulsion zone ending at a distance r_r, followed by the alignment zone ending at r_o, and the attraction zone ending at r_a, i.e. $r_r < r_o < r_a$. The direction of the social forces is computed by

$$\hat{f}_{i,rep} = \mathcal{N}\left(\sum_{j \in \mathbb{S}_{i,r}} -\hat{r}_{ji}\right) \tag{10.4a}$$

$$\hat{f}_{i,alg} = \mathcal{N}\left(\sum_{j \in \mathbb{S}_{i,o}} \vec{v}_{ji}\right) \tag{10.4b}$$

$$\hat{f}_{i,att} = \mathcal{N}\left(\sum_{j \in \mathbb{S}_{i,a}} \hat{r}_{ji}\right) \tag{10.4c}$$

with $\hat{x} = \vec{x}/|x|$ defining an unit vector and $\vec{x}_{ji} = \vec{x}_j - \vec{x}_i$ as the difference between the vectors of fish j and i. $\mathbb{S}_{i,x}$ is the set of fish in zone x of fish i and $\mathcal{N}(\vec{x}) = \hat{x}$ is a normalization operator. I assume Voronoi interactions because they provide a reasonable approximation to visual networks (Strandburg-Peshkin et al., 2013) and can be efficiently computed. Therefore, the sets $\mathbb{S}_{i,x}$ with $x \in [r, o, a]$ are composed only of Voronoi neighbors of fish i. Note that the alignment force is the sum of the velocity difference vectors $\vec{v}_{ji}$. Thus, the focal fish i experiences the strongest alignment with neighbors whose velocity vectors differ the most from its own.

If neighbors of fish i occupy different zones simultaneously, the following rules apply:

- if $\mathbb{S}_r \neq \emptyset$: $\hat{f}_{soc} = \hat{f}_r$ (repulsion dominated)
- if $\mathbb{S}_r = \emptyset$ and $\mathbb{S}_o \neq \emptyset$ and $\mathbb{S}_a \neq \emptyset$: $\hat{f}_{soc} = \mathcal{N}(|\mathbb{S}_a|\hat{f}_a + |\mathbb{S}_o|\hat{f}_o)$ (weighted average)

10.1.3 Environmental force

In the absence of a predator the environmental force is modeled as a random force vector. The randomness is a consequence of the assumption that shoaling fish misinterpret environmental noise (e.g. water reflections, water perturbations, sounds) as threats since there are no "real" threats in the experiment. In the presence of a threatening agent the environmental force is modeled as a simple repulsion of the shoaling fish from the simulated predator or fishing agents

$$\hat{f}_{env} = \begin{cases} [cos(\varphi), sin(\varphi)] \text{ with } \varphi \sim U(-\Pi, \Pi) & \text{no predator present, } \mathbb{S}_{i,p} = \emptyset \\ \mathcal{N}\left(\sum_{p \in \mathbb{S}_{i,p}} -\hat{r}_{pi}\right) & \text{otherwise} \end{cases} \tag{10.5}$$

with $U(a, b)$ being an uniform distribution with a and b as lower and upper bounds, and $\mathbb{S}_{i,p}$ being the set of predator or fishing agents detected by a shoaling fish i. The probability that a shoaling fish detects a predator or fishing agent is probabilistic

$$P_{detect} = \min\left(\frac{2}{1 + r_{pi}/r_f}, 1\right) \ . \tag{10.6}$$

r_f is the detection distance and is for all simulations $r_f = 7\ cm$. A detection distance of $r_f = 7\ cm$ is reasonable because fish should be able to detect a predator when they are likely to be captured $r_f \geq r_{capture} = 5\ cm$ but the distance should also be close to $r_{capture}$, otherwise fish would respond too often to non-dangerous cues. Additionally, the decay with distance represents the reduced visibility in water with larger distances.

10.1.4 Wall-avoidance mechanism

The model is designed to be as close to the experimental setting as possible. Since in the shoaling experiment (Sect. 9.4) a circular tank was used, the simulated fish have a circular boundary. The introduction of parameters describing the repulsion force from the wall is avoided by the fish following a parameter free wall-avoidance mechanism. It is based on predicting before each burst the future position $\vec{r}(t_1)$ of the shoaling fish at the next burst (at t_{nb}, which stands for "time to next burst"), plus some extra time (at time $t_1 = t_{nb} + t_b$). The extra time t_b is necessary because the agent can not prevent collision with the finite burst force if it is at the next burst inside the tank but directly at, or very close to the wall. The length of the extra time is set to the burst time t_b, because if the agent coasts (i.e. uses no force) and does not collide with the wall, any burst force is sufficient to prevent collision if the agent bursts instead. If the predicted position $\vec{r}(t_1)$ is outside of the tank, the force direction is adapted until $\vec{r}(t_1)$ is inside the tank. Thereby, the smallest possible change in force direction is used. This mechanism prevents the agents from colliding with the boundary by modifying their intended movement direction as little as possible.

To predict the future position $\vec{r}(t_1)$ with $t_1 = t_{nb} + t_b$ the solution of the coupled differential equation for $\vec{v}(t)$ and $\vec{r}(t)$ is used (Eq. 10.1). Since the x and y components of $\vec{v}(t)$ and $\vec{r}(t)$ are

not coupled with each other, the solutions are computed for each component separately:

$$v(t) = (v_0 - F/\beta)e^{-\beta t} + F/\beta \tag{10.7a}$$

$$r(t) = \frac{1}{\beta}(v_0 - \frac{F}{\beta})(1 - e^{-\beta t}) + \frac{F\ t}{\beta}\ . \tag{10.7b}$$

I omitted the index r_x or r_y for the terms $r,\ F,\ v_0$ and v for simplicity. The solutions above assume a constant force F. Therefore, first, the position and velocity after the burst are predicted and then, second, the position after the coast.

10.2 Model comparison

The presented model is not the first attempt to imitate the interactions of fish with a burst-and-coast swimming behavior (Calovi et al., 2018; Harpaz et al., 2017). The two existing studies used zebrafish (*Danio rerio*, Harpaz et al., 2017) and Rummy-nose tetra (*Hemigrammus rhodostornus*, Calovi et al., 2018) as an experimental reference species. Both models, as well as the presented burst-coast model, have an active phase in which the agent accelerates and a passive phase in which they decelerates due to friction. Here, I shortly explain both models and compare them with the model presented in this chapter.

In the study by Harpaz et al. (2017), from now on referred to as Harpaz model, the agents respond solely to the velocity of agents. Thereby, the focal fish adopts the velocity term of a neighboring agent weighted according to its spatio-temporal bin it belongs to. There are in total 108 spatio-temporal bins (6 angular sectors x 6 ring-sectors x 3 past time sectors). For each bin exists another weight that defines the interaction with the wall. This sums up to 216 parameters. Please note that this is the upper limit of parameters. Some weights are set to zero by penalizing high weights in the fitting process with the Lasso least square regression. However, more than half the fitted weights have a value clearly different from zero. Importantly, the modeled agents switch between active and passive phase, e.g. burst and coast phase, such that the error to an experimental trajectory is minimized. Thus, it is not a generative model (it does not generate trajectories fully by its own), but this could be easily implemented by a switch rate.

The parameters in the model by Calovi et al. (2018), from now on referred to as Calovi model, are not weights of spatio-temporal bins but parameters in analytical functions. It is an elegant approach, because symmetry arguments are used to define the general mathematical structure of the functions that regulate the strength of the attraction and alignment between agents and the repulsion from the wall. The burst duration and speed are drawn from distributions fitted to experimental observations. Thus, the model has a stochastic character because it is based on random draws from distributions and because of the noise on the heading-direction. In total about 18 parameters need to be fitted to describe the full model (5 alignments para. + 5 attraction para. + 2 wall para. + 4 distribution para. + 1 friction para + 1 noise para.). The model shows a very good agreement with the data and interestingly supports that fish align with each other. The fitting of the model is based on experimental data of single fish and fish pairs in a tank. From the single fish experiments the strength of the noise on the heading change during bursts and the interaction with the wall is estimated. They approximate the width of the Gaussian distribution of heading noise to be around 20°. The fish pair trajectories are used to estimate the social force parameters.

The presented burst-coast model, in the following referred to as zonal model, has aspects of both models: it has discrete interaction zones as in the Harpaz model, however only 3 bins in total (compared to 108), and considers repulsion attraction and alignment as in the Calovi model. From a mathematical point of view the Calovi model is preferable, because the analytical dependence of the forces is based on reasonable symmetry arguments and it reproduces the experimental data not only in terms of mean-values but also in terms of distributions. However, the motivation of this study is to demonstrate that an attention trade-off mechanism can explain differences between selection lines. For this purpose, I have chosen the probably most simple model, 8 parameters vs 18 (Calovi) and 216 (Harpaz), that is in principle able to demonstrate the mechanism. Furthermore,

the shoaling experiment was not designed to fit the parameter as in the Calovi model (using single fish and fish pair experiments). From the shoaling experiment only trajectories of groups of eight fish are available. Even if tracks of fish pairs existed, it is questionable to just upscale the system from single pair interactions (Gautrais et al., 2012). However, one apparent discrepancy between the Calovi model and the zonal model is that the former limits the noise on the change in heading during bursts to draws from a Gaussian distribution with width 20°. The zonal model does not limit the noise, i.e. agents follow environmental cues by drawing the force direction from an uniform distribution spanning the complete angular space (see Eq. 10.5).

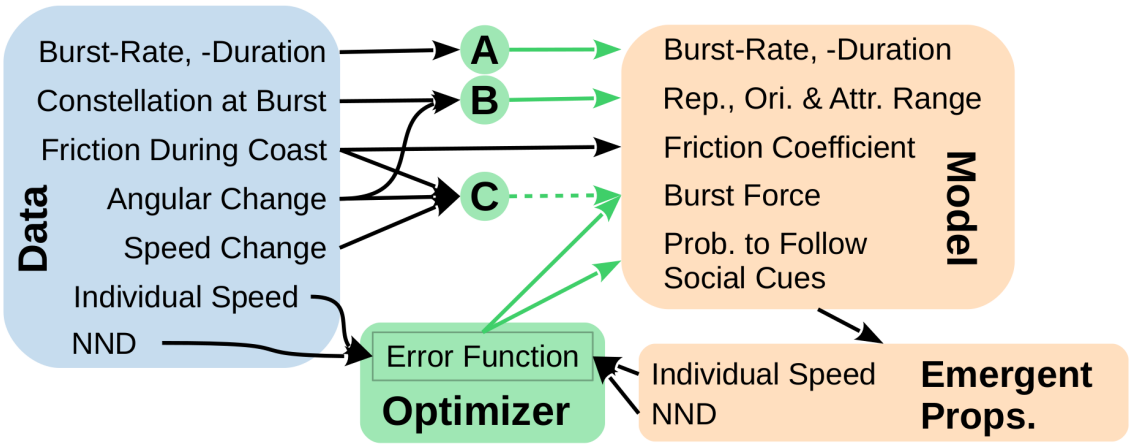

Figure 11.1: Burst-coast parameter setting overview Light green boxes/circles represent optimizer that minimize an error-function by adjusting the model parameters. However, the optimizer marked by A-C do not simulate the full model but only isolated aspects of it. In contrast, in the optimizer that defines the burst force and the probability to follow social cues (green box), the full model is simulated because emergent properties (nearest neighbor distance and the mean individual speed) are compared to the experimental measures. The central green circle marked with an "C" represents the optimizer that deduces the burst force parameter. However, the estimation of the burst force F is likely to be too low (as explained in Sect. 11.1.2) and therefore is estimated by the lower optimizer.

11 Model representations of selection lines

In this chapter the model is fit to each of the three selection lines. Thus, each selection line will be represented by a unique set of parameters. The chapter is split in four sections.
First, most parameters (6 of 8) are estimated without the need to simulate the model. Their values can be directly or indirectly inferred from experimental data (upper three arrows connecting experiment with model in Fig. 11.1).
Second, the remaining parameters (2 of 8) can not be inferred from data, and therefore I simulate the model and compare the nearest neighbor distance and the average individual speed with the experimental measures (lower green box in Fig. 11.1). The model representations of the selection lines support the proposed mechanistic explanation. The representations fail to reproduce the experimental results with a common probability to follow social cues (P_{soc}) for all selection lines. Only if this probability (which is closely related to vigilance) varies in a way that is consistent with the risk-taking experiments (Sect. 9.3), the experimental cohesion differences are reproduced.
Third, the details of the simulated model representations are analyzed and alternative explanations of the cohesion differences are explored.
Forth and finally, the results are discussed.

11.1 Simulation-free parameter estimation

In the model, an agent can either be in the burst or coast phase (Sect. 10.1.1). The trajectory is split in burst and coast periods to estimate the friction coefficient β, the burst rate γ and duration t_b, the burst force f_b and the social ranges. Burst and coast phases are characterized by an increase and decrease in the speed of the shoaling fish, respectively.

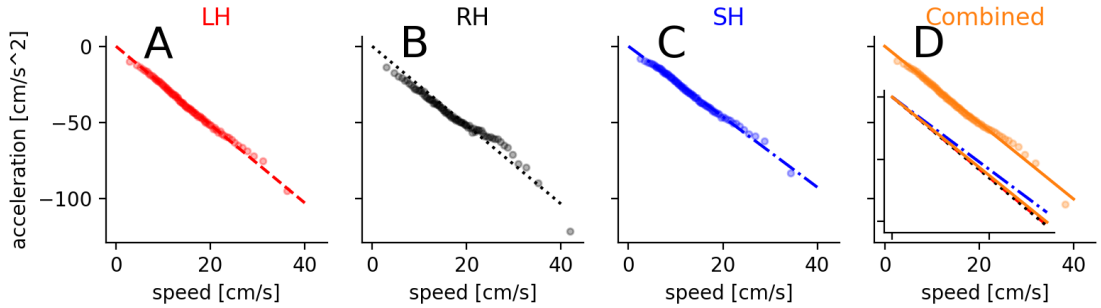

Figure 11.2: Estimation of the friction coefficient β from experimental data of the large harvested (LH, **A**), random harvested (RH, **B**), small harvested (SH, **C**) and of all selection lines combined (**D**). Dots represent mean of bins with varying width such that each bin contains an equal amount of data. The lines are linear fits with the interception at zero and the negative friction coefficient as slope. Inset **D**: all linear fits are shown together with the LH, RH, SH and combined fits are represented by red dashed, black dotted, blue dash dotted, orange solid lines, respectively. For the model representations of the selection lines the slope of the combined selection lines was used to estimate the friction coefficient $-\beta = -2.51$.

11.1.1 Friction coefficient

In the coast phase the only parameter that defines the change in velocity is the friction coefficient β (Eq. 10.2). Thus, β is estimated by the negative slope of the linear regression of the acceleration on the current speed of the fish during coast-phases (Fig. 11.2). Because the friction force is by definition zero at zero velocity, I fixed the interception point to zero and therefore computed the least-square solution of the linear equation $\frac{dv}{dt} = -bv$. The linear dependence of the acceleration is clearly valid for all three selection lines (Fig. 11.2 A-C).

The estimated friction coefficients are almost identical, only the SH line (inset Fig. 11.2 D) has a slightly smaller friction coefficient. This can be explained by (i) the friction depending on the cross section of the fish and (ii) differences in mass between the selection lines, which are assumed to be the zero. The cross-section was not measured. However, it is reasonable to assume that it increases with increasing body length, which is largest for the SH line. Therefore, the friction should naively be expected to be largest for the SH line assuming it is based on Stokes law. That is not the case. However, in Eq. 10.2 the friction coefficient is actually an effective friction coefficient, i.e. the real friction coefficient divided by the mass of the agent. Therefore, the friction coefficient decreases with increasing mass, which would explain the lower effective friction coefficient of the SH line. Note that if the shape is held constant but the volume of the fish increases, the Stokes radius is likely to increase less than the mass. This is evolutionary beneficial, because otherwise heavier fish need disproportional more force to keep the same speed and would spent more energy to avoid predators.

However, in order to keep the number of parameters low, I assume the same friction coefficient $\beta = 2.51\ 1/s$ for the model representations of all selection lines. In addition, the burst force also influences mainly the individual average speed. Therefore, possible differences in the friction coefficient can be represented by modifications of the burst force strength F. The next section shows that the estimation of the burst force is problematic and consequently is estimated instead by fitting the individual average speed (and the nearest neighbor distance) to the experimental measures.

11.1.2 Burst force strength

The burst force F in heading direction $f_{b,s}$ can be estimated by adding to the change in speed $\frac{dv}{dt}$ the friction coefficient times the current velocity:

$$f_{b,s} = \frac{dv}{dt} + \beta \cdot v \ . \tag{11.1}$$

This results in the mean burst force in heading direction $f_{b,s} = 95cm/s^2$. However, according to Eq. 10.2 the total magnitude of the burst force is defined as

$$f_b = \sqrt{f_{b,s}^2 + f_{b,\varphi}^2} \tag{11.2}$$

with $f_{b,\varphi}$ as force in direction perpendicular to the current heading direction. The latter is (derived from Eq. 10.2c)

$$f_{b,\varphi} = \frac{d\varphi}{dt} \cdot v \ . \tag{11.3}$$

Note, that f_b instead of the symbol for the model parameter F is used, because I expect F to be larger than f_b since the latter is estimated from forces only during the burst-phase. However, the turning force often started prior to the acceleration in velocity direction and reaches its maximum at the start of the burst (Fig. 9.2C, D). This suggests that the actual force F needed to mimic the characteristic zigzag-like swimming of zebrafish is larger than f_b. For the different selection lines I estimated $\langle f_{b,LH} \rangle = 121.2cm/s^2$, $\langle f_{b,RH} \rangle = 135.2cm/s^2$, $\langle f_{b,SH} \rangle = 102.3cm/s^2$. Instead of setting $F = f_b$, I used its largest mean to set the boundaries of the search-space for F in the optimizer that finds the burst force F and probability to follow social cues P_{soc} that reproduce best the mean individual swimming speed and the nearest neighbor distance.

Modifications by adapted burst-durations

From Eqs. 11.1, 11.3 it becomes clear that the actual burst duration might be underestimated. The burst phases are marked by positive "naive" accelerations in speed, i.e. the fish is declared as bursting if $dv/dt > 0$. Coast phases are analogously marked by the negative "naive" acceleration. From the coast phases the friction coefficient β is estimated. This friction coefficient acts however all the time and the burst and coast phases need in principle a re-evaluation by using Eqs. 11.1, 11.3. However, this causes also a re-evaluation of the friction coefficient that in turn makes the re-estimation of burst and coast phases necessary. This has not been done and probably the friction coefficient and the burst and coast phases converge after repeating this procedure, which need verification. It is another good reason why the burst forces should be set by fitting emergent variables of the model to its experimentally observed analogs.

11.1.3 Rate and duration of bursts

The average burst duration $\langle t_b \rangle$ and rate $\langle \gamma \rangle$ are estimated by the mean length and frequency of burst periods. For the large-, random- and small-harvested line the burst durations are $\langle t_b \rangle = [0.117, 0.123, 0.123]s$ and the burst rates are $\langle \gamma \rangle = [0.35, 0.44, 0.38]$, respectively. In the model, a new burst period can start during an already ongoing burst and therefore prolong the measured burst duration and decrease the rate, as illustrated in Fig. 11.3. Consequently, the average burst duration and rate is different from the model parameters t_b and γ. I approximated them by creating a binary time series in which at each time-step dt a burst-event happens with probability $\gamma \cdot dt$, which raises the acceleration from zero to F for a duration of t_b. I simulated this process, computed the average values $\langle t_b \rangle_N$, $\langle \gamma \rangle_N$ (as seen in Fig. 11.3) and selected the γ and t_b that minimize the summed square differences to the experimental observed averages for each selection line:

$$err_{t_b,\gamma} = \left(\frac{\langle t_b \rangle - \langle t_b \rangle_M}{\sigma_{t_b}} \right)^2 + \left(\frac{\langle \gamma \rangle - \langle \gamma \rangle_M}{\sigma_\gamma} \right)^2 \ . \tag{11.4}$$

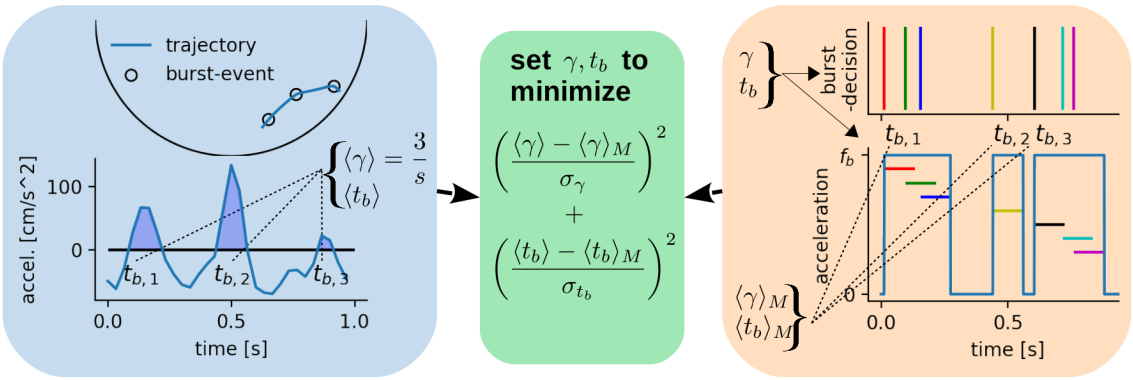

Figure 11.3: Parameter setting of the burst rate γ and duration t_b. Blue box: computation of the experimental averages $\langle\gamma\rangle$, $\langle t_b\rangle$ from an example trajectory. Orange box: model averages $\langle\gamma\rangle_M$, $\langle t_b\rangle_M$ are computed for model parameters γ, t_b by generating a time series in which overlapping burst are considered at one burst. Green box: optimizer minimizes the difference between the averages based on data ($\langle\gamma\rangle$, $\langle t_b\rangle$) and on model-simulations ($\langle\gamma\rangle_M$, $\langle t_b\rangle_M$ orange box). It selects the parameters of the smallest error in a two dimensional parameter scan. The error is defined by the sum of normalized square differences of γ and t_b between experiment and model. For the normalization the experimentally observed standard deviations σ_{t_b}, σ_γ are used.

The normalization is done with the experimental standard deviation of the respective measure, i.e. σ_{t_b}, σ_γ. For each selection line the squared difference of γ are shown in Fig. 11.4 A-C and of t_b in Fig. 11.4 E-G. The model parameters of the burst rate γ are set at about twice the value as the experimental observed average. The experimentally observed burst rate differences are conserved, i.e. $\langle\gamma\rangle_{RH} > \langle\gamma\rangle_{LH} > \langle\gamma\rangle_{SH}$. The burst duration did differ between the selection lines only in the order of milliseconds. The parameters are reported in Tab. 11.2.

11.1.4 Ranges of social zones

The ranges of the social-interactions zones (r_r, r_o, r_a) are set to minimize the angle between the predicted and the actual direction after a burst (Fig.11.5). The computation of the predicted direction is based on the relative positions of neighbors of the focal, bursting fish at the start of the burst. From this neighbor constellation the direction of the social force is computed with Eq.10.4a. Only burst-event are considered where the bursting fish had a minimal distance of three body length to the tank wall. How well a given choice of interaction zone ranges explains the data was estimated by the mean angle difference between its predicted and the actual direction after the burst. To find the parameter-choice for r_r, r_o and r_a that minimizes this angle difference we, first, ran two different optimizer (dual-annealing: python/scipy implementation of (Xiang et al., 2013), differential-evolution: python/scipy implementation of (Storn and Price, 1997)), which gave for each selection line similar parameters. For all selection lines the width of the orientation zone was below 2 millimeters. This suggests that the best solution favors no alignment at all. To verify this, two-dimensional parameter scans around the optimal parameter setting were performed (see Sect. B.2.1 and Fig. B.4). For the scan in which the repulsion r_r and orientation range r_o are varied, the angular difference is lowest if $r_o = r_r$, i.e. a two-zone model without orientation zone explains best the data (Fig. B.4B, E, H). Consequently, I set the $r_o = r_r$ and repeated a parameter-scan around the optima of the selection lines (Fig. 11.6). From this scan I extracted the ranges reported in Tab.11.2. Interestingly, the ranges of the large-harvested line are smaller than those of the random-harvested line. The ranges of the RH lines are again smaller than those of the small-harvested line. Therefore, the ranges could reflect body-length differences in between the selection lines.

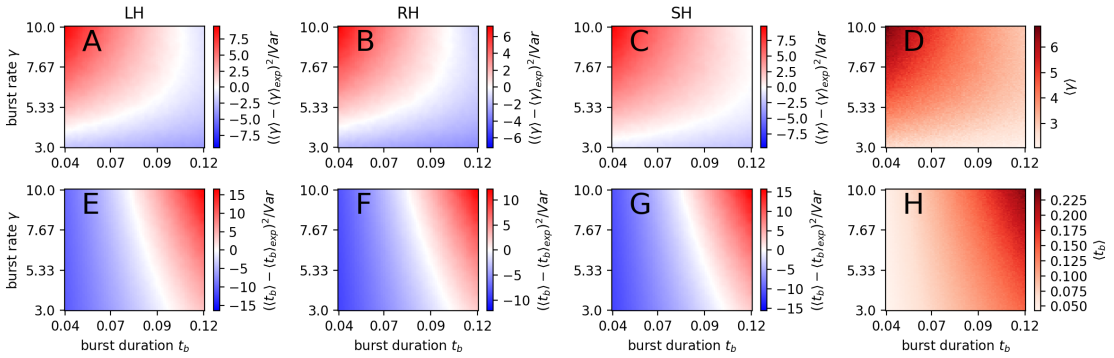

Figure 11.4: Model parameter estimation of burst duration t_b and burst rate γ (parameters of an acceleration-time-series generating model). The acceleration-based-average is reported for burst rate $\langle\gamma\rangle$ (**A-D**) and burst duration (**E-H**) $\langle t_b\rangle$. The normalized squared difference between the averaged values of the simulation $\langle x\rangle$ and the experiment $\langle x\rangle_{exp}$ are shown for large-harvested line (LH; **A, E**), random-harvested line (RH; **B, F**) and small-harvested line (SH;**C, G**). The normalization was done by dividing the squared differences by the experimental variance of the burst rate (**A-C**) or of the burst duration (**E-G**). The average values of the simulation are shown in (**D, H**).

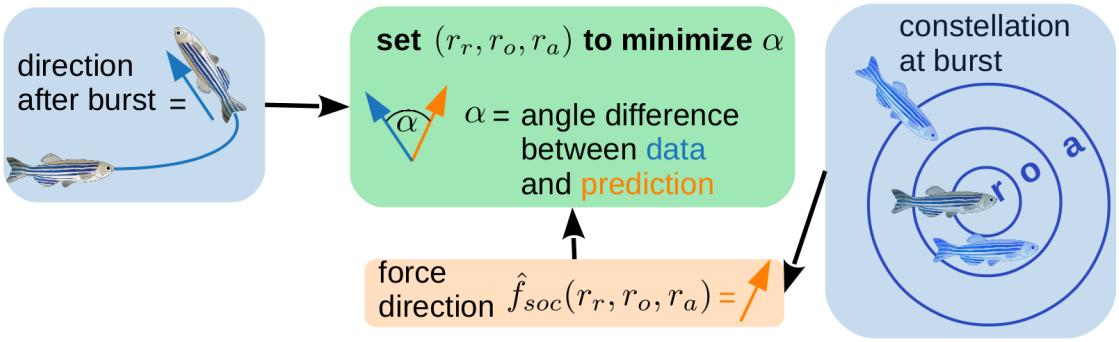

Figure 11.5: Sketch of how the range of repulsion r_r, orientation r_o, and attraction r_a are set by minimizing the angle difference α between the model-predicted force direction $\hat{f}_soc$ and the actual heading direction after a burst. The force direction for specific r_r, r_o, r_a is predicted from the relative positions and velocities of the neighbors of the bursting individual at the start of the burst, i.e. from the neighbor constellation at burst.

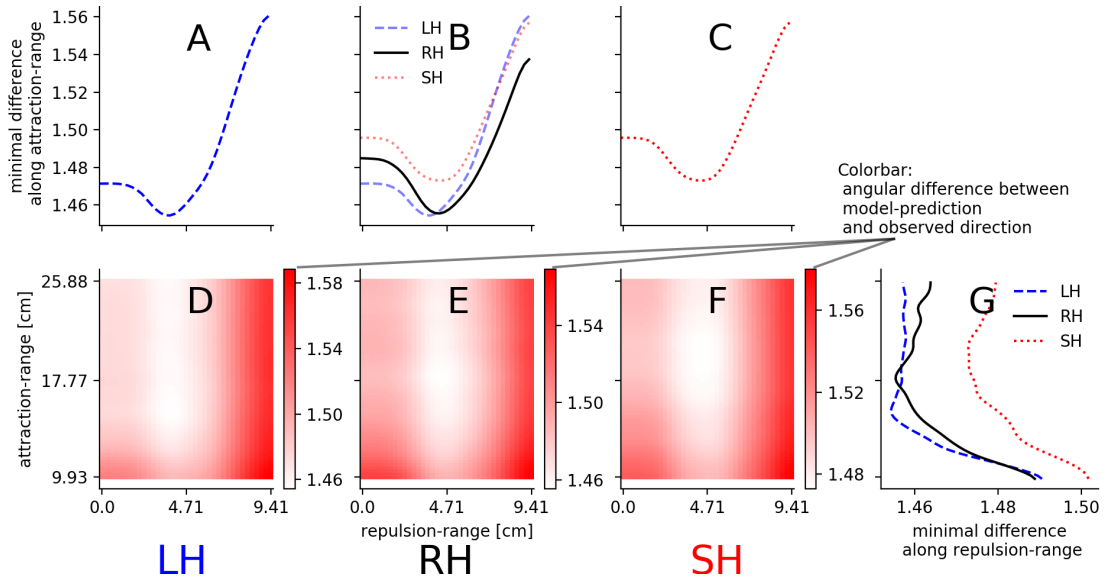

Figure 11.6: Angle difference between predicted and actual direction after burst is color-coded (**D-F**) or its minimal value along one axis direction is shown (**A-C, G**). Here, in contrast to the scans in Fig.B.4, no orientation-zone exists, i.e. $r_o = r_r$. **A-C**: minimal angle difference along the attraction-range is shown for burst-constellation data of the large- (**A**), small-harvest (**C**) and for all lines (**B**). **D-F**: angular difference for different parameters of repulsion and attraction range for the large-(**D**), random-(**E**) and small-harvested line(**F**). **G**: minimal difference for a specific attraction-range for all selection lines.

11.2 Parameter setting by fitting model-emergent measures

11.2.1 Method

In the above section most parameters are estimated from experimental data. This section explains how the remaining parameters, burst force F and the probability to respond to social cues P_{soc}, are set by fitting model-emergent measures to the experimental counterparts 11.1. To fit the model to the data I applied the Covariance Matrix Adaptation Evolution Strategy (CMA-ES; Hansen et al., 2019; Hansen and Ostermeier, 1996). This optimizer is a good choice if the fitness landscape is multi-modal, the search space dimension is between 4 and 100 and no gradient is known. The optimizer minimizes the error function

$$E(\mathbb{P}_{LH}, \mathbb{P}_{RH}, \mathbb{P}_{SH}) = \frac{Err(\mathbb{P}_{LH}) + Err(\mathbb{P}_{RH}) + Err(\mathbb{P}_{SH})}{3} \tag{11.5}$$

with $\mathbb{P}_{\mathrm{SL}}$ as the set of parameter of a selection line (SL). Eq. 11.5 averages over the selection line errors. The error of a specific selection lines compares the measured nearest neighbor distance NND and average individual speed v to the experimental data by computing the sum of the squared differences standardized by their experimental standard deviation:

$$Err(\mathbb{P}_{\mathrm{SL}}) = \left(\frac{\langle \mathrm{NND} \rangle_{\mathrm{SL}} - \langle \mathrm{NND} \rangle_{M,\mathrm{SL}}}{\sigma_{\mathrm{NND,\ SL}}} \right)^2 + \left(\frac{(\langle v \rangle_{\mathrm{SL}} - \langle v \rangle_{M,\mathrm{SL}}}{\sigma_{v,\mathrm{SL}}} \right)^2 . \tag{11.6}$$

The index M in $\langle \mathrm{x} \rangle_{M,\mathrm{SL}}$ marks that x is computed from model simulations. $\langle \mathrm{x} \rangle_{\mathrm{SL}}$ and $\sigma_{x,\mathrm{SL}}$ are the experimental mean and standard deviation of the measure x. The experimental means of NND and v used for the error function are listed in Tab. 11.1.

At each generation the parameter sets are updated by the CMA-ES. Note that the parameter sets differ in burst-duration, -rate, repulsion-, attraction-range and probability to respond to

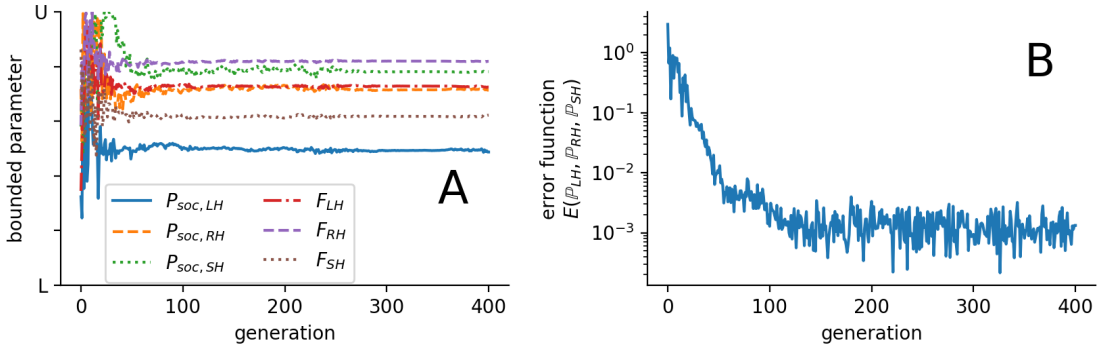

Figure 11.7: Example of optimization run with the parameters that result in the lowest error in the current generation (**A**) and the corresponding error (**B**). The parameters are normalized according to their respective lower and upper boundary (see Tab. 11.1), which are marked on the y-axis with L and U (**A**). Different colors and line styles mark the probability to respond to social cues P_{soc} of the large (solid blue), random (dashed orange) and small (dotted green) harvested line and the burst forces F of the large (dash dotted red), random (dashed violet) and small (dotted brown) harvested line.

social cues in between the selection lines. For each selection line the burst force F and the probability to respond to social cues needs setting. Consequently, the search-space is 6-dimensional ($P_{soc,LH}$, $P_{soc,RH}$, $P_{soc,SH}, F_{LH}$, F_{RH}, F_{SH}) as highlighted in Tab. 11.2.

The search-space of the algorithm for the three different parameters is limited by setting the boundaries that are listed in Tab. 11.1. For the probability to follow social cues, a minimum attention to social and environmental cues was ensured by setting the boundaries 0.05 above and below the theoretical possible boundaries of zero and one. I expect the burst force to be larger than the experimentally estimated measure, as discussed in section 11.1.2. Thus, its boundaries are half and twice the mean burst force of the random harvested line.

To ensure that the resulting minimum is not a local minimum the optimization is repeated from different initial parameter settings. The initial parameter were selected from a two-dimensional grid with 2 grid-points and therefore $2^2 = 4$ different initial parameter settings. Note that the actual search-space is 6-dimensional ($P_{soc,SH}$, $P_{soc,RH}$, ...). By setting P_{soc} and F for the different selection line initially equal, the number of initial settings is reduced from $2^6 = 64$ to 4. The initial standard deviations in the covariance matrix are a quarter of the search space range. This is sufficient for an overlap of the explored parameter space of the two different initializations in each search dimension.

An example optimization run over 400 generations is shown in Fig. 11.7.

11.2.2 Results

The optimization outcomes of the 4 different initialization are shown sorted by their final error according to Eq. 11.5 in Fig. B.5. The smallest error is $E(\mathbb{P}_{LH}, \mathbb{P}_{RH}, \mathbb{P}_{SH}) \approx 0.0009$. Thus, the difference between model and experimental measures is on average 0.02 standard deviations. In Fig. 11.8A the parameters F and P_{env} are shown that resulted in the smallest difference to the experimental measures. The dashed lines show the different initialization that resulted in a slightly larger error. The parameters vary minimal in between different initializations. This means that the optimizer found the global minimum.

The fitted burst forces of the selection lines obey the same inequality as the experimentally estimated burst forces (see Sect. 11.1.2), i.e. $F_{RH} > F_{LH} > F_{SH}$. As predicted, in Sect. 11.1.2, are the fitted burst forces much larger as the experimental estimates. The probability to follow environmental cues, i.e. $P_{env} = 1 - P_{soc}$, is largest for the LH line and the RH line is only

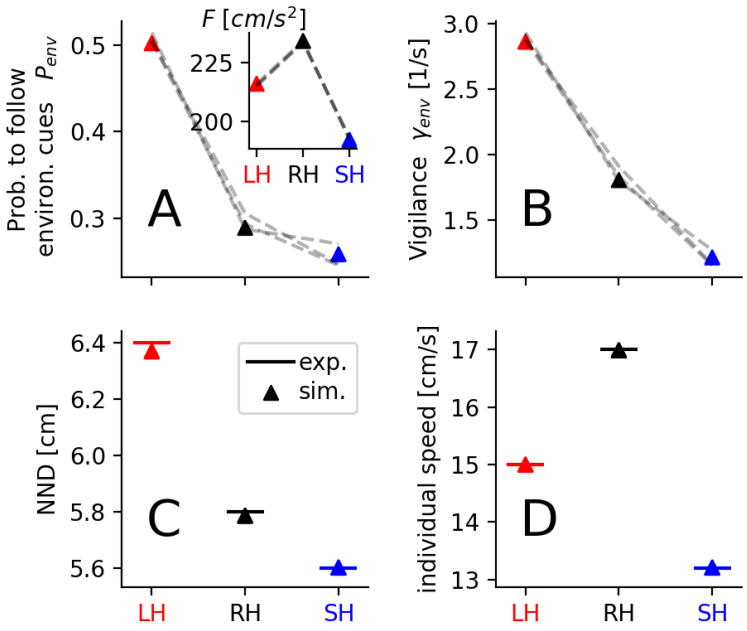

Figure 11.8: Fitted model representation of the different selection lines. In order to resemble the experimental observations the probability to follow environmental cues (**A**) and the burst force (inset **A**) were set for each selection line. The vigilance γ_{env} is the product of P_{env} and the burst rate γ (**B**). The average nearest neighbor distance (NND, **C**) and the average individual speed (**D**) are emergent properties of the model and were used to quantify how well the parameters (P_{env}, γ) reproduce the experimental observations. Triangles represent the model parameters or the simulation outcomes of the parameter set with the best match to the experiment. Dashed lines (**A, B**) represent a parameter set of a different initialization, and therefore show the robustness of the best matching parameter set (triangles). The horizontal solid lines (**C, D**) represent the experimental values. The selection lines are indicated by different colors (red=LH: large-harvested; black=RH: random-harvested; blue=SH: small-harvested).

about three percent more likely to follow environmental cues than the SH line. The closely related vigilance (product of burst rate and P_{env}) was for the large-harvested line greater than control, and vice versa for the small-harvested line (11.8B). Since the burst rate γ of RH line is about 40 percent faster than γ of the SH line, the difference in vigilance between RH and SH is larger than the difference in P_{env}.

In Fig. 11.8C, D the emergent model measures NND and individual average speed are compared to the experimental counterparts. It visualizes the extremely small error that I reported earlier as $E(\mathbb{P}_{LH}, \mathbb{P}_{RH}, \mathbb{P}_{SH}) \approx 0.0009$. Note that in earlier versions of the model different parameters have been fit to the experimental measures with a comparable error (e.g. P_{soc} for each selection line and common F and a common enhanced turning parameter α that not present in the current model). However, different initializations of earlier models did not result in the same parameters. This suggests that only local minima have been found.

Note that the use of the CMA-ES is not needed if the selection lines do not share the parameters that are optimized, which is the case. Each selection line has its own F and P_{soc} and contributes equally to the total error (Eq. 11.5). Thus, the minimum of the total error must be composed of the minimal errors (Eq. 11.6) of the isolated selection lines. In principle, the minimum error of a simple two-dimensional (F, P_{soc}) parameter scan for each selection line is the preferable method. It does not rely on a complicated optimizer and the results are easily verifiable. However, historically the selection lines shared parameters that coupled the errors of the selection lines. Furthermore,

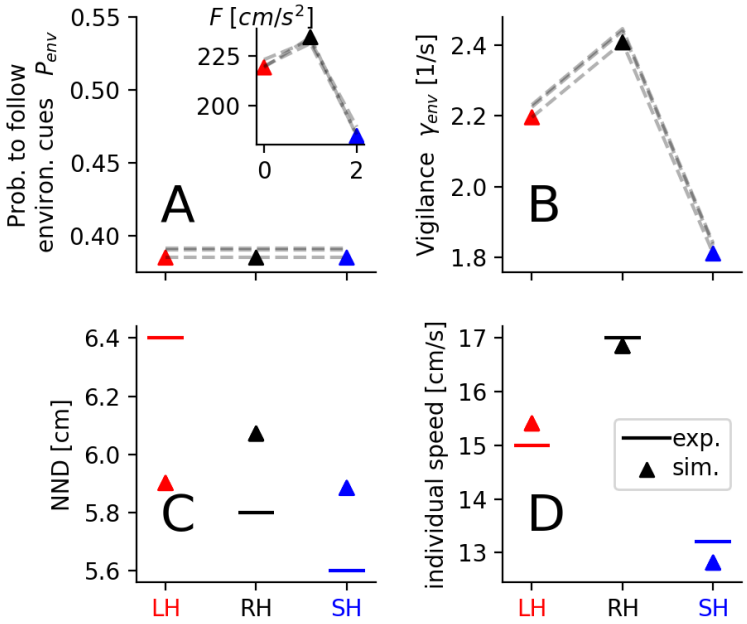

Figure 11.9: Fitted model representation of the different selection lines with same P_{env}. Note that here, in contrast to Fig. 11.8, the probability to follow environmental cues is the same for all selection lines. For detailed description see Fig. 11.8.

to identify the explanatory variable of the cohesion pattern, I will enforce in the next subsection that the selection lines have a common P_{soc}. This is not feasible with a two-dimensional parameter scan (a four dimensional scan would be necessary) but the here used CMA-ES is also in this case appropriate.

11.2.3 The explanatory variable for the cohesion pattern

The differences in vigilance (Fig. 11.8B) are qualitative similar to differences in risk taking behavior (Sect. 9.3). Therefore, a possible mechanistic explanation for the observed differences in cohesion is that the more risk-taking individuals of the small-harvested line were less vigilant, which causes them to respond less frequently to environmental cues, instead of social cues. Consequently, the cohesion of the SH line increased (vice versa for the large-harvested line).

To substantiate this explanation I re-estimated the parameters F, P_{env} by enforcing that all selection line shared the same P_{env}. Without the ability to differ in P_{env} (Fig. 11.9A), the differences in vigilance are solely due to the differences in burst rate γ, which in turn do not resemble the differences in risk taking behavior anymore (Fig. 9.3). This was to be expected, and the mechanistic explanation does not hold anymore. Importantly, the model representations are now unable to reproduce the experimentally observed cohesion pattern (Fig. 11.9C). Therefore, the explanatory variable in the proposed model representation for the cohesion pattern are the differences in P_{env} or equivalently in vigilance γ_{env}.

It is possible that the cohesion pattern vanishes because the number of parameters is reduced. If the burst force parameter F is the same for all selection lines and P_{env} is still allowed to vary between the selection lines, the cohesion pattern is quantitatively reproduced (Fig. B.6C). In addition, the individual speed of the selection line model representations qualitatively agrees with the experiments (Fig. B.6D). Importantly, the estimated P_{env} still varies between selection lines as the risk taking behavior (Fig. B.6A). It supports the mechanistic explanation. The common burst force F was optimized to $F = 216.4 cm/s^2$. In this case the number of model parameters (3 x P_{env}

+ 1 x F) is the same as if P_{env} is shared between selection lines (3 x F + 1 x P_{env}). Therefore, not the reduced number of parameters disables the model to reproduce the cohesion pattern but the inability to vary in a specific parameter, i.e. in P_{env} that is proportional to vigilance γ_{env}.

This claim only holds if I assume that the model and how its parameters have been fitted to experimental data produce a fair representation of the unknown real fish behavior. Remember, all parameters, but the friction coefficient β, do vary between the selection lines. Thus, each parameter difference in between selection lines could explain the cohesion pattern, but if P_{env} does not vary between the lines, the experimental cohesion pattern is not reproduced by the model. In the next section I explicitly investigate the effect of other parameters on the model, assuming that parameter estimations where inaccurate.

11.3 Model exploration

In this section I (i) explore alternative mechanisms that could produce the cohesion pattern and (ii) compare properties of the model that have not been fitted explicitly to the data.

11.3.1 Cohesion sensitivity on parameter variations

Sect. 11.2.3 suggested that the explanatory variable is P_{env}. I assumed that the simulation-free parameter estimation (Sect. 11.1) was accurate and produced a fair model representation. Here, I show that most parameters can in principle affect the cohesion (nearest neighbor distance NND) and other measures as the polarization $\Phi = 1/N \mid \sum_i \vec{v}_i/v_i \mid$ and mean individual speed.

The NND is influenced by the ranges of the social zones in an expected manner, i.e. cohesion decreases with larger repulsion r_r and/or smaller attraction r_a(Fig. 11.10A). An increase in orientation range r_o effectively reduces the attraction zone and therefore causes a loss in cohesion. The parameters that are expected to strongly affect the individual speed (t_b, γ, F, β) show a loss in cohesion with an increase in speed (compare Fig. 11.10B with E). The only exception is the burst rate γ. It has has no effect on NND because a higher rate not only increases the speed but also the response to social cues that regulate the NND. With an increase in social responsiveness P_{soc}, the cohesion increases (Fig. 11.10C). This increase is weakened if the social response incorporates more alignment (compare dashed with solid line in Fig. 11.10C). Again, this is due to the decrease of attraction zone width.

The average individual speed increases with larger orientation range r_o but is unaffected by changes in r_r and r_a (Fig. 11.10D). The speed associated parameters modify the individual speed in the expected way, i.e. an increase in t_b, γ, F or a decrease in β increase the speed (Fig. 11.10E). The probability to follow social cues P_{soc} only causes a speed increase if the alignment zone is increased compared to the standard parameter setting (Fig. 11.10F).

The polarization naturally increases with an increase in orientation r_o, which also happens if the repulsion range r_r is reduced (Fig. 11.10G). It is, however, unaffected by all other parameters (Fig. 11.10G-I) but by P_{soc}, which increases the polarization only if the orientation range is increased (Fig. 11.10G-I). Without an increase of r_o an increase in P_{soc} decreases the polarization, because only the attraction and repulsion forces act, which in turn interrupt the spontaneous alignment.

The cohesion, measured via the NND, responds in principle to almost all parameter variations (Fig. 11.10A-C), the individual speed is strongly affected by four out of eight parameters and the polarization effectively only by one, the orientation range r_o (Fig. 11.10G-I). Note that I explored the eight dimensional parameter space only in an one dimensional fashion around the fitted parameters of the LH line model representation. An exception is the additional increase of the orientation zone to $r_o = 10$. This made P_{soc} more influential, because the fitted model representations had no orientation zone before, i.e. $r_o = r_r$. The one dimensional parameter explorations behave as expected and illustrate that deviations from the parameter estimations in Sects. 11.1, 11.2 could also explain the cohesion pattern.

optimization parameter	symbol	optimization boundaries	unit
burst force	F	$[67.6,\ 270.4]$ $(=[\langle f_b\rangle_{RH}/2,\ 2\langle f_b\rangle_{RH}\ ]$	cm/s^2
prob. of social burst	P_{soc}	$[0.05,\ 0.95]$	1
error function defining measures	**symbol**	**experimental mean and STD**	**unit**
individual speed	v	LH: 15 ± 2.8 RH: 17 ± 2.5 SH: 13.2 ± 2.7	cm/s
nearest neighbor dist.	NND	LH: 6.4 ± 1 RH: 5.8 ± 1.1 SH: 5.6 ± 0.9	cm

Table 11.1: Optimization search space boundaries and error function relevant measures. The first part of the table lists the upper and lower boundaries in which the optimizer looks for the minimum of the error function Eq. 11.5, i.e. they define the search space. The lower and upper boundary of the burst force F are given by half and twice the mean of the experimental estimated burst force $\langle f_b\rangle_{RH}$ of the RH line. The boundaries for the probability to follow social cues ensure a minimum of social and environmental responsiveness. The second part of the table lists the measures defining the error function Eq. 11.5 that is minimized by the optimizer.

model parameter	symbol	estimated value	estimation method	unit
repulsion range	r_r	LH: 3.92 RH: 4.44 SH: 4.71	data+model	cm
alignment range	r_o	r_r	data+model	cm
attraction range	r_a	LH: 15.84 RH: 18.45 LH: 19.5	data+model	cm
burst force	F	LH: 215.1 RH: 234.4 SH: 193	optimizer	cm/s^2
prob. of social burst	P_{soc}	LH: 0.5 RH: 0.71 SH: 0.74	optimizer	1
burst duration	t_b	LH: 0.089 RH: 0.091 SH: 0.097	data+model	s
burst rate	γ	LH: 5.7 RH: 6.3 SH: 4.7	data+model	$1/s$
friction coefficient	β	2.51	data	$1/s$

Table 11.2: Model parameters of the burst-coast model. The estimation methods are colored according to the colors used in Fig. 11.1. Each selection line has its specific parameter, which is indicated by the abbreviations LH: large-harvested; RH: random-harvested and SH: small-harvested. An exception is the friction coefficient β, which is the same for all lines,

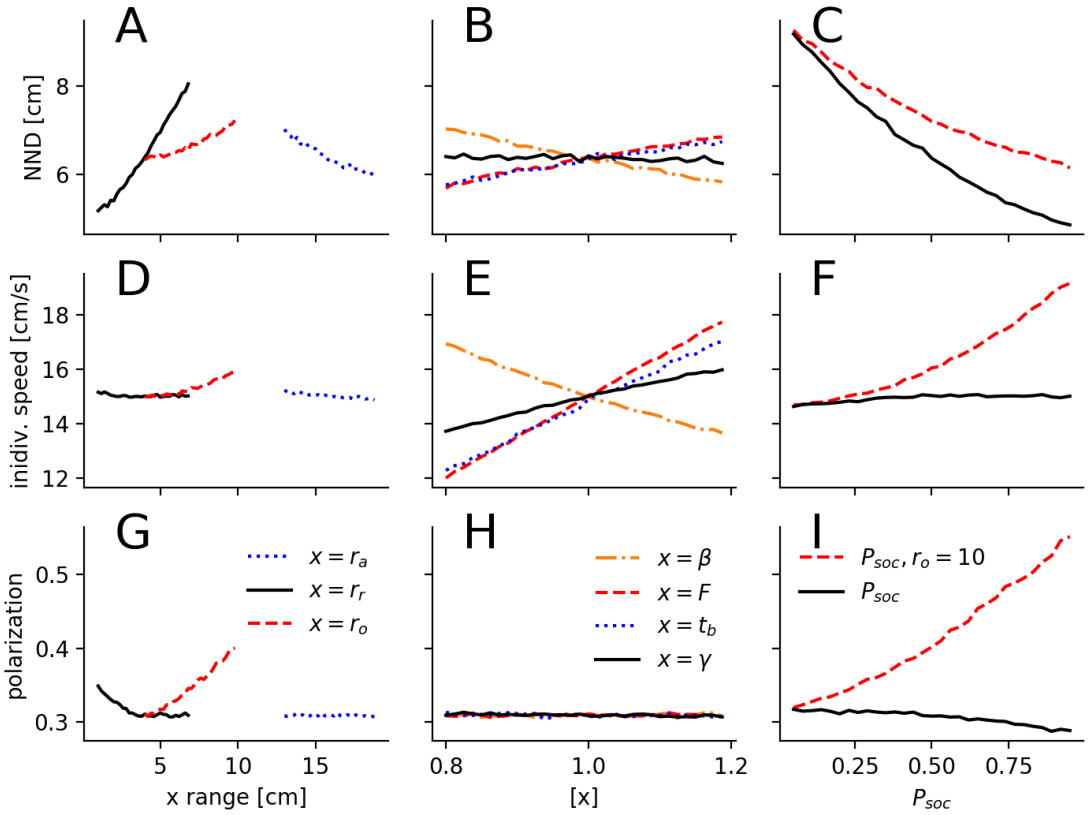

Figure 11.10: Parameter exploration in the burst-coast model. Exploration on how the mean nearest neighbor distance NND (**A-C**), the mean individual speed (**D-F**) and the polarization (**G-I**) are influenced by changes of model parameters. First column (**A,D,G**): ranges of the social zones are varied. The black solid, red dashed and blue dotted line represent repulsion- r_r, orientation- r_o and attraction range r_a, respectively. Second column (**B,E,H**): parameters with a strong impact on speed are varied. The black solid, red dashed, blue dotted and orange dash-dotted line represent burst rate γ, burst force F, burst duration t_b and friction coefficient β, respectively. The unit of the x-axis is the respective parameter value from the model representation of the LH line, e.g. $0.8x = 0.8 \cdot 2.511/s$ for x representing the friction coefficient β. Third column (**C,F,I**): variation of the probability to respond to social cues P_{soc} with standard parameter (black solid line) and with the orientation range set to $r_o = 10cm$ (red dashed line). In all panels the unchanged parameters are the ones from the model representation of the large harvested line (Tab. 11.2).

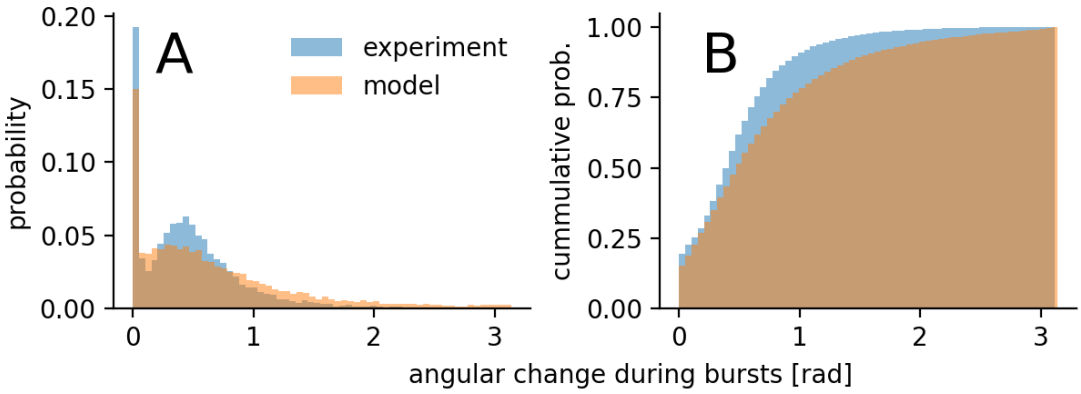

Figure 11.11: Comparison of angular change during bursts. The angular changes observed in experimental data are compared with model simulations via their probability distributions (**A**) and cumulative probability distributions(**B**). Blue bars represent experimental measures of all RH lines and the orange bars represent the simulations results of the model representation of the RH line (Tab. 11.2).

11.3.2 Model-data comparison

In Sects. 11.1, 11.2 the model parameters were fitted to the experimental data. Here I investigate if unfitted measures of the model resemble the experimental observations.

Calovi et al. (2018) introduced a burst-coast model where the random force direction was drawn from a Gaussian distribution with a width of 20 degree. The proposed burst-coast model draws it from an uniform distribution spanning the whole circle. Therefore, it is interesting to investigate how the angular change during a burst in the model compares to the data (Fig. 11.11). The distribution of angular changes in the model shows a distinct peak at zero and a rather smooth transition with a long tail up to the maximal angular change, i.e. π. The experiment shows as well the distinct peak at 0 radians but, unlike in the model, a second peak at about $\pi/8$ and no long tail, i.e. the maximal angular change is at about $\pi/2$. There are at least three possible explanations. First, real fish follow a direction that represents a compromise between its current heading direction and the force direction. Second, real fish have a reduced turning due to angular friction, which is neglected in the model. Or, third, the draw from an uniform distribution is unrealistic. However, the third explanation does not solve the problem because agents also respond to social forces which can in principle also point in any direction.

The mean of the individual speed distribution was used as an emergent variable to fit F and P_{soc}. However, also the variance, which was not fitted, and the general shape of the distribution, resembling a Poisson distribution, are recovered by model representation (Fig. 11.12A). Note that the maxima of the distributions do not overlap, because the model distribution has a similar but larger variance (experiment: $\sigma_s = 7.4 cm/s$, model: $\sigma_s = 9.3 cm/s$). The model variance would shrink if (i) agents, as mentioned above for the difference in angular changes, would follow a compromise direction between heading and force direction and (ii) the burst force would not be discrete but continuous. The condition (i) makes the burst force more parallel to the current heading direction, which would push the speed distribution to larger values. The condition (ii) reduces the burst force, because, as explained further below, a larger fixed force is necessary to fit the smaller average of a varying burst force.

The polarization, which was not fitted, agrees well with the experiments in mean (model: $\langle \Phi_M \rangle \approx 0.3$, experiment: $\langle \Phi_M \rangle \approx 0.4$) and variance (model: $\sigma_{\Phi,M} \approx 0.15$, experiment: $\sigma_{\Phi,M} \approx 0.19$) (Fig. 11.12B). Again, if the fish would follow a compromise direction, it would reduce their noise and consequently increase the polarization of the model, resulting in a closer fit. Another explanation why the experimental polarization is larger than the model analog is that real fish are attracted to

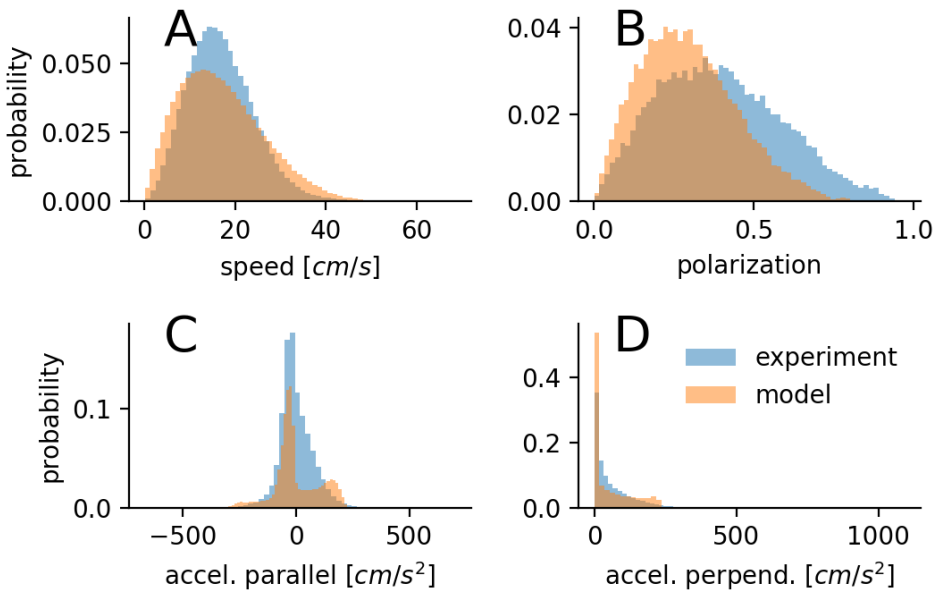

Figure 11.12: Comparison of unfitted model characteristics to data. The distribution of the individual speed (**A**), the polarization (**B**), the acceleration parallel (**C**) and perpendicular (**D**) to the current heading direction. Blue bars represent experimental measures of all RH lines and the orange bars represent the simulations results of the model representation of the RH line (Tab. 11.2).

the tank wall. It is also observable from the video that real fish tend to favor positions close to the walls because the wall represents the only cover in the tank. The simulated agents do not favor wall position. However, if they would be attracted to the wall, about half the possible movement directions would be excluded because of the wall avoidance mechanism. In addition, movements away from the tank would be suppressed because of the wall attraction. Both mechanism could result in a larger polarization.

The strongest discrepancy between model and data can be seen in the distribution of the acceleration parallel (Fig. 11.12C) and perpendicular (Fig. 11.12C) to the current heading direction. While the distribution is mono modal in the experiment, it is bimodal for the model. This is in line with the proposed model-modification to allow the burst-force to vary. It also explains why the fitted burst force $F_{RH} \approx 234 cm/s^2$ is almost by a factor two larger than the experimental estimate $f_{b,RH} \approx 135 cm/s^2$, and why the tail of the model speed distribution is longer compared to experiment. In order to fit a medium average acceleration a much larger fixed acceleration needs to be chosen.

In summary, the model resembles the measures of the experiment that have not been fitted. Possibly, only a few adjustments are necessary, i.e. a varying burst force and a directional compromise between current heading and intended force, to render the model more realistic. However, the model was never designed to capture the experiment in such detail. Thus, instead of introducing unnecessary complexity to an idealized model an alternative model approach might be more promising. To clarify how the burst force should vary is also a non trivial task. A possible solution is to assign different burst strength to the repulsion, orientation and attraction zone. Since individuals interact in most cases with multiple neighbors, the combination of the forces can result in the necessary variation. Still, (i) it is complicated to estimate the strength of those forces, but an analog approach to Calovi et al. (2018) seems promising, and (ii) the environmental burst would still be fixed. To solve problem (ii) the strength of the burst force could be drawn from a distribution (again this distribution needs to be estimated).

11.4 Summary and discussion

I fitted a distinct parameter set to each selection line. Thereby, only the friction coefficient β was the same across lines. The same β was not fully supported by data (Fig. 11.2) but it does impact mainly the speed, as does the burst force F (see Fig. 11.10E). Since the burst force is fitted to the emergent variables NND and individual speed, it can incorporate the differences in β. During the parameter exploration in Sect. 11.3.1 it became apparent that also the NND distance is affected by β. However, the burst force F affects the NND in the opposite way, as for the speed. Therefore, the additional influence on NND is unproblematic.

I found that the model represents the data best if the orientation zone is neglected (see Sect. 11.1.4). This is in agreement with force-fitting of golden shiners (*Notemigonus crysoleucas*; Katz et al., 2011) and mosquito fish (*Gambusia holbrooki*; Herbert-Read et al., 2011) where no signs of alignment where detected but it contradicts the force-fitting in Rummy-nose tetra (*Hemigrammus rhodostornus*; Calovi et al., 2018) and in *Danio rerio* (Zienkiewicz et al., 2018). Note that Zienkiewicz et al. (2018) found an alignment between pairs of zebrafish. A possible explanation is that we analyzed groups of 8 fish and that alignment is dominating only in pair interactions and weakens for larger groups (as shown in experimental data and model fits in *Kuhlia mugil* Gautrais et al., 2012). Thus, it is possible that all studies correctly identified the social mechanisms specific to each species and group size. However, each study approach varies in its assumptions and a re-estimation of the social forces for each species using different approaches would be very informative. Thus, the lack of orientation in zebrafish seems possible but should be taken with caution. The force estimation might for example change if the social zones are not exclusive but additive (Calovi et al., 2018; Couzin et al., 2005, e.g. as in).

The missing orientation zone causes the probability to follow social cues P_{soc} to mostly affect the cohesion and not the other emergent variables (individual speed, polarization). However, this is not why P_{soc} is identified as the explanatory variable and by no means a necessary condition. The argumentation is rather that all parameters can be fairly approximated for each selection line representation but P_{soc} and the burst force F. And from these two only variations in P_{soc} allow to reproduce the cohesion pattern. This is because F influences the individual speed much stronger than the NND. Therefore, F is used by the optimizer to adapt the speed pattern instead of the cohesion pattern. Surely, if I would allow a parameter with a strong impact on the cohesion (as the repulsion range r_r) to be estimated by the emergent variable comparing optimizer, the P_{soc} would not be the only explanation.

The Sect. 11.3.2 indicated that the proposed simplistic burst-coast model is able to reproduce experimental measures that where not used in the model fit. However, only up to a certain degree. Namely, compared to the experiment, the model has a wider variance in speed, a bimodality of the acceleration distribution and larger angular changes during bursts. The simplest intuitive explanation is that the fixed and constant burst force is unrealistic and a varying force should be taken into account.

Given the proposed burst-coast model, the tendency to follow social cues P_{soc} is the explanatory variable. It supports the proposed mechanistic explanation of the cohesion pattern. Importantly, this explanatory variable links the risk taking experiments consistently via the vigilance to the cohesion pattern.

12 Predictions of context dependent adaptation

The model representations of the selection lines enables us to predict how each lines perform in natural predators and fishing context. Of particular interesting is whether the large harvested selection line, associated with strong commercial fishing pressure, experiences greater exploitation by natural predators. If so, even after a decrease in commercial fishing the population could shrink, because they are an easier target for natural predators.

In this chapter, I first describe the three exploitation scenarios: (i) natural predation, (ii) uninformed fishing agent (angler in a boat or spearfishing) and (iii) multiple informed fishing agents aligned in a line. In the second part, the results of the scenarios are compared.

12.1 Scenario descriptions

I used the agent-based model representations of the three selection lines to investigate whether the size-selective harvest could impact the ability of the shoals to evade a natural predator and different fishing gear. $N = 30$ shoaling fish are simulated in a box of size $L = 100\ cm$ with periodic boundary conditions. Three scenarios are considered: (i) natural predation - including the confusion effect (Milinski, 2010a,b) - by a single mobile predator following the closest fish; (ii) fishing by a single agent without information about the position of the shoal (random search, similar to a fisher on a boat without an echo sounder); (iii) fishing by multiple aligned agents moving on a straight line towards the center of mass of the shoal (e.g., commercial trawling with information on the shoal position).

In contrast to the simulation with only shoaling fish, now if the fish detect a predator, the environmental force models a repulsion away from the predator (Eq. 10.5). Please recall that the fish always detect the predator/fishing agent if it is closer than the detection distance $r_f = 7\ cm \approx 3BL$ (Eq. 10.6). For larger distances the detection probability decays linearly until it equals zero at $r = 35\ cm \approx 14BL$. For details see Sect. 10.1.3.

All additional parameters for the three scenarios are explained below (summarized in Tab. 12.1).

12.1.1 Natural predator

The predator moves directly to the closest fish with $v_{net} = 20,\ 25,\ 30,\ 35\ cm/s$, which is larger than the average speed of shoaling fish ($\langle v \rangle \approx 15\ cm/s$) because predators are usually larger than prey and therefore can swim faster (Domenici, 2001; Howland, 1974). Most predators attack a specific fish and therefore need to focus on it before attacking. The so-called confusion effect (Landeau and Terborgh, 1986), the disruption of the predator focus by a large number of individuals who are difficult to distinguish by phenotype and movement, is believed to be one key benefit of group living (Krause and Ruxton, 2002). We model the probability of a prey to be successfully captured if it is closer to the predator than $r_{capture} = 5cm$ within a small time window $[t, t + \delta t]$ as

$$p_{success,i}(t, \delta t) = p_{confuse}(t) \cdot \gamma_a \delta t \tag{12.1}$$

Here $\gamma_a = 1/s$ is a base predator attack rate and $p_{confuse}(t)$ represents the confusion effect that modulates the attack rate. The confusion term decreases $p_{success}$ with increasing number of perceived prey N_{sensed} in a sigmoidal fashion

$$p_{confuse} = \frac{\tanh(-(N_{sensed} - N_{conf})) + 1}{2}\ . \tag{12.2}$$

$N_{conf} = 4$ is the number of sensed shoaling fish at which $p_{confuse} = 0.5$ (Landeau and Terborgh, 1986), and a fish is sensed if it is closer than $r_{sense} = 4r_{capture}$.

12.1.2 Fishing agents

In contrast to the natural predator, the fishing agent's capture rate is not decreased by the confusion effect. This allows to simplify the simulations by assuming that fishing agents always capture a shoaling fish if it is closer than $r_{capture} = 5\ cm$. This assumption is equivalent with a maximal capture rate and corresponds to an encounter-based capture mechanism typical for many fisheries (Alós et al., 2012; Sbragaglia et al., 2018). However, in some forms of fishery the maximal capture is not a reasonable assumption, for example in angling where each successful catch is followed by a preparation time (pull in the fish, prepare new lure). Still, it should not alter the results if a smaller capture rate is assumed, but most likely only prolong the simulation time.

Fishing gears were simulated by varying the speed of the fishing agents ($v_{net} = 7.5,\ 15,\ 22.5\ cm/s$). Compared to the average swimming speed of zebrafish ($\langle v \rangle \approx 15 cm/s$), these velocities correspond to slower, equally fast and faster fishing agents.

Single fisher

In the single fishing agent scenario, the agent performs a random search with constant speed v_{net}. The individual fishing agent is supposed to have a constant persistence length. It's angular change is given by

$$\frac{d\varphi}{dt} = \hat{\sigma}_\varphi \xi(t)\ . \tag{12.3}$$

$\xi(t) \propto \mathbb{N}(0,1)$ is Gaussian white noise with zero mean and variance of one. $\hat{\sigma} = \hat{\sigma}(v_{net})$ is the angular noise strength whose exact form and dependence is derived below.

Keeping the persistence length constant is equivalent with keeping the variance in angle constant after the individual fishing agent traveled a path of length $= l$. The time needed to travel this path length is $t_l = l/v_f$ and the variance in angle after traveling the path is

$$\begin{aligned} Var(\varphi)(l, v_{net}) &= \langle \varphi(l, v_{net})^2 \rangle && \text{(12.4a)} \\ &= \left\langle \left(\int_0^{t_l} \hat{\sigma}_\varphi \xi(t) dt \right)^2 \right\rangle = \hat{\sigma}_\varphi^2 \int_0^{t_l} \int_0^{t_l} \langle \xi(t)\xi(t') \rangle dt dt' && \text{(12.4b)} \\ &= \hat{\sigma}_\varphi^2 \int_0^{t_l} \int_0^{t_l} \delta_{t,t'} dt dt' = \hat{\sigma}_\varphi^2 \int_0^{t_l} dt && \text{(12.4c)} \\ &= \hat{\sigma}_\varphi^2 t_l = \hat{\sigma}_\varphi^2 \frac{l}{v_{net}}\ . && \text{(12.4d)} \end{aligned}$$

I exploited the uncorrelated nature of Gaussian white noise, e.g. $\langle \xi(t)\xi(t') \rangle = \delta_{t,t'}$, with $\delta_{t,t'}$ as the Kronecker-delta. If I set $\hat{\sigma}_\varphi = \sigma_\varphi \cdot \sqrt{v_{net}}$, the variance in angle after traveling a path of length l is independent of the speed:

$$\begin{aligned} Var(\varphi)(l) &= \hat{\sigma}_\varphi^2 \frac{l}{v_{net}} && \text{(12.5a)} \\ &= \sigma_\varphi^2 \cdot l\ . && \text{(12.5b)} \end{aligned}$$

Multiple fisher

In the multiple aligned fishing agents scenario, the agents ($N_f = 50$; red) are aligned on a line spanning $L/4$. After traveling for a distance $L/2$, the fishing agent array is recreated at a distance of $L/4$ away from the center of mass of the shoal and restarts its movement in its direction.

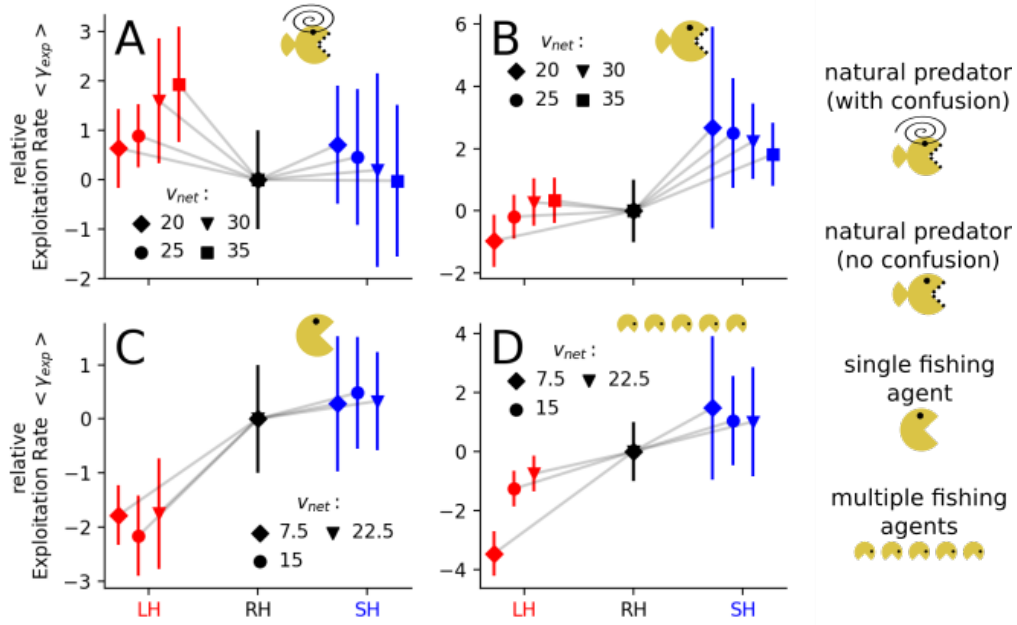

Figure 12.1: The exploitation rate, i.e. prey captured per time unit, was computed for a natural predator with (**A**) and without confusion (**B**), for a single fishing agent (**C**) and for multiple fishing agents distributed on a line (**D**). The natural predator moves with $v_{net} = [20,\ 25,\ 30,\ 35]\ cm/s$ indicated by rhombus, circle, triangle and rectangle, respectively (**A, B**). The fishing agents move with $v_{net} = [7.5,\ 15,\ 22.5]\ cm/s$ indicated by rhombus, circle and triangle, respectively(**C, D**). For comparison, the shoaling fish have a mean speed of about $\langle v \rangle \approx 15\ cm/s$. Shown are relative mean exploitation rates, i.e. the mean exploitation rates were reduced by the mean and divided by the standard deviation of the random-harvested line. The error bars indicate relative the standard deviation. 400 simulations of $N = 30$ shoaling fish in a box of size $L = 100\ cm$ with periodic boundary conditions were used. The colors red, black and blue correspond to the large- (LH), random- (RH) and small-harvested (SH) selection line, respectively.

12.2 Context dependent exploitation of selection lines

In this section, the derived model representations of each selection line are simulated in the three exploitation scenarios. Since the natural predator mainly differs because of the confusion effect, the natural predator simulation is repeated without the confusion effect.

First, the exploitation rate γ_{exp} is compared in between selection lines. It is the number of captured individuals per time unit (second). Remember that the LH line is the least cohesive line, which means that the natural predator should be less confused and thus attack more often compared to the other lines. The simulations of the natural predator with confusion support this expectation (Fig. 12.1A). If the confusion effect is not present, the LH line is no longer the fastest exploited (Fig. 12.1B). In contrast, I expect that the exploitation rate of the SH line remains unchanged with respect to the RH line. Both lines are similar in NND, thus they are expected to behave in the same way with respect to the confusion effect. However, while the RH line is exploited about as fast as the SH line by a confused predator (Fig. 12.1A), it is slower exploited than the SH line for a non-confused predator (Fig. 12.1B). This is the case because of the larger mean individual speed of the RH line and because the predator is only in average faster than individual prey agents. Thus, the faster prey of the RH line has larger peak velocities (at the end of a burst phase) and therefore can escape farther than the SH line during the burst phase. Note that in the coast phase, the predator is decreasing the distance again. This explanation is in line with a decreasing exploitation difference between RH and SH lines: the predator swims at the

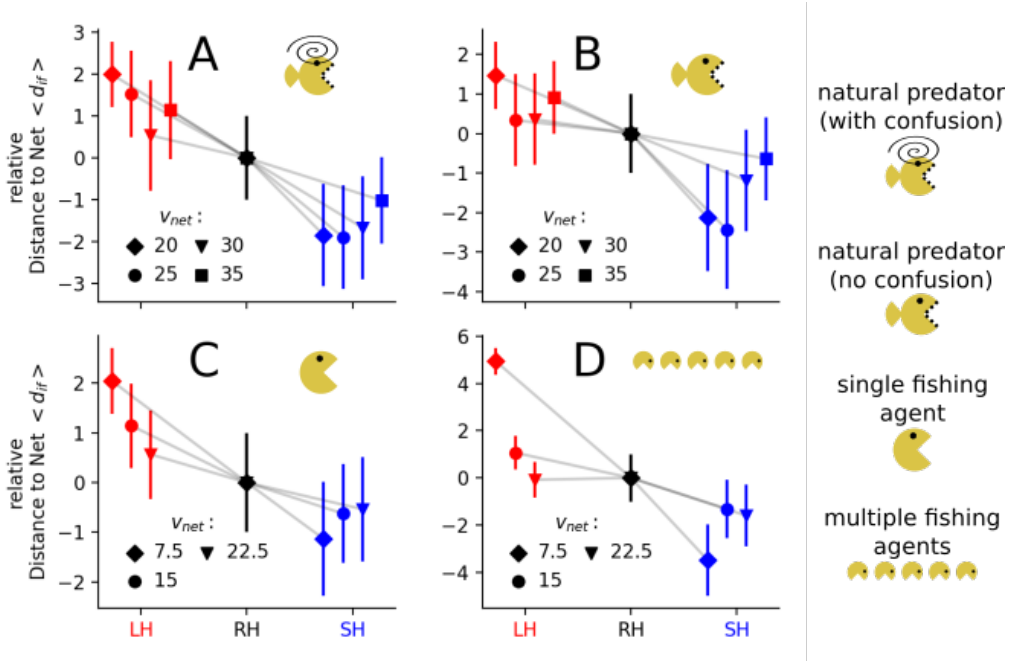

Figure 12.2: The mean distance between prey and the natural predator with (**A**) and without confusion (**B**). The mean distance between prey and a single fishing agent (**C**) and multiple fishing agents distributed on a line (**D**). For further details see caption of Fig. 12.1.

prey agent's peak-velocity. Therefore, the small differences in individual speed are not sufficient anymore to grant the RH line a benefit compared to the SH line.

The single fishing scenario and the multiple fishing scenario show that the LH line is exploited slowest and the SH line fastest for all different speeds of the fishing agents (Fig. 12.1C, D). This is because the LH line is most vigilant and SH line is least vigilant line, i.e the LH line is more likely to detect the fishing agents and more likely to escape. Note that the RH line is not exploited less due to its higher individual speed compared to the other lines for the fishing scenarios. This is due to the encounter based deterministic capture process, i.e. a sufficient close prey agent is always captured by the fishing agents but has only an increased chance to get caught by a natural predator. Is that an artifact or does a realistic interpretation exists? It could represent the reduced mobility of fish that come in contact with a fishing net or with the spear of speargun.

An alternative measure is the mean distance a shoaling agent keeps from fishing agents:

$$\langle d_{if} \rangle = \frac{1}{N} \sum_{i=1}^{N} \min_{f \in \mathbb{S}_f} d_{if} \tag{12.6}$$

with $d_{if} = |\vec{r}_{if}|$ as the distance between shoaling fish i and a fishing agent f. The minimum function yields the distance to the closest fishing agent. This measure helps to explain why the RH line is in the natural predator scenario with and without confusion always among the two least exploited lines. If the explanation given above (pursued agents of the RH line keep a larger distance because of their higher average speed) is true, the average distance to the predator/fishing agent should not be affected by the higher speed of the RH line. Because only the closest agent increases its distance, but the other agents should increase their distance according to the selection line specific P_{env}. This prediction is observed for the distance measure d_{if}, i.e. the LH line has in each scenario the largest d_{if} and the SH line the smallest d_{if} (Fig. 12.2). Thus, the differences in d_{if} correspond to differences in vigilance between the selection lines (Fig. 11.8B)

12.3 Summary and discussion

In this chapter, I used the model representation of selection lines to predict how strong they are exploited in different contexts relative to each other. The LH line was exploited at the slowest rate in both fishing scenarios but fastest exploited in the natural predation scenario. This allows the interpretation that the behavior modifications of a fish species under strong recreational or commercial fishing (LH line, increased mortality for large individuals is associated with fishing with nets of large mesh-size or releasing only small fish after a successful catch) are less exploited by commercial fishing with respect to previous generations (previous generations represented by the RH line). But in the same time these fish lack a natural defense mechanism, i.e. they are less cohesive, which allows natural predator to exploit them faster. Consequently, a decrease in commercial fishing would not necessarily lead to a recovery of the fish species because natural predators keep exploiting them at high rates. This could lead to a hysteresis effect, i.e. only a much larger decrease in commercial fishing would restore fish populations to a pre-fishing abundance.

The recent study by Guerra et al. (2020) found this hysteresis effect as well, but due to a mechanism solely based on group size. They assume that fish have (i) the ability to estimate their own group size, (ii) a preferred group size and a changing behavior if this preferred group size is crossed and (iii) that the preferred group size is coded in genes, i.e. it is inherited by the offspring. The group size dependent behavior is the following: if the group is smaller than preferred, individuals tend to initiate to merge with another group, and if the group is larger than preferred, individuals tend to initiate a group fragmentation. The evolution acts on the preferred group size that renders this assumption vital. Fishery pressure is modeled by decreasing the fitness of individuals that are part of larger groups. In contrast, fish that are more often part of small groups are less fit under strong natural predation. With this proposed mechanism by Guerra et al. (2020), the LH line would also be less exploited by commercial fishing and stronger by natural predators because their cohesion is weaker and therefore their final group size is smaller.

Note that in the predictions we used groups of $N = 30$ individuals of the model representations that have been fitted to experimental data of groups of eight individuals. It is not clear how social behavior changes with group size in zebrafish, but it has been reported for other species (Gautrais et al., 2012). In golden shiner the social forces estimated in larger groups are very distinct from the groups of 2 individuals, especially it is not clear if the forces are additive or need to be averaged (Katz et al., 2011). Thus, the predictions for the different scenarios should be taken with caution.

It is obvious that zebrafish are a poor model species to make predictions and imply impact of commercial fishing on pelagic shoaling fish. Still, if the results are relevant for other shoaling pelagic fish, it is of importance for the fishing industry because 13 of the 25 most fished species are shoaling (FAO, 2018, Tab. 3).

scenario related parameter	symbol	value	scenario relevance	unit
flee range[a]	r_f	7	all	cm
fisher/predator capture range	$r_{capture}$	5	all	cm
predator sensing range	r_{sense}	$4r_{capture} = 20$	natural pred.	cm
confusion number	N_{conf}	4	natural pred.	cm
attack rate	γ_a	1	natural pred.	$1/s$
angular noise strength	σ_φ	0.2	single fisher	$1/s^{1/2}$

Table 12.1: Parameters for predator/fishery scenarios. The column "scenario relevance" specifies for which scenario the parameter is needed. [a]: Note that the flee range influences the probability of a prey detecting a predator/fisher (Eq. 10.6) which is 0.5 at a distance of $21cm$ given $r_f = 7cm$. Therefore, the range of the prey to detect a predator is comparable to the predators sensing range.

13 Summary and discussion

This part of the thesis had a strong link to experimental data. It inspired a modeling approach that introduced a trade-off between social and environmental information. Importantly, the trade-off persists even in the absence of environment information. The new model aspect is inspired by unexpected cohesion differences between differently size harvested selection lines, given their documented risk-taking behavior. I fitted the burst-coast model to experimental data by choosing a simpler model than already existing studies (Calovi et al., 2018; Harpaz et al., 2017, only 8 parameters compared to 18 and 216). The model representations of the selection line revealed that the proposed explanation for the differences in cohesion is very reasonable. To be precise, it turned out that the differences in vigilance, i.e. the rate of bursts based on environmental information, is causing the cohesion differences and that the vigilance differences are qualitatively similar to the risk taking behavior differences. I also demonstrated that in principle differences between selection lines in other parameters than P_{env} could cause the cohesion pattern, given that their fit due to model assumption or technical errors is incorrect. Finally, I demonstrated that the selection lines behave different to each other in different contexts. Of specific interest is the LH line (representing a fish population under commercial fishing pressure) which is stronger exploited by natural predators but less strong by commercial fisheries. This could mean that an over fished population is more vulnerable to natural predators and even further decreases if commercial fishing is reduced (possible hysteresis effect).

The burst-coast-agent-based model linked risk-taking behavior with shoal cohesion. Furthermore, It helped to explain the size selective mortality induced changes in collective behavior by assuming the existence of a trade-off between social and environmental information(Rahmani et al., 2020). The proposed model is similar to previous work in terms of splitting the movement of shoaling fish into an active and passive phase (Bode et al., 2010; Calovi et al., 2018; Harpaz et al., 2017). However, it provides a novel perspective about collective movements of fish shoals by linking risk perception and movement decisions. In fact, most previous models of collective behavior only accounted for social information (Bode et al., 2010; Calovi et al., 2018; Couzin et al., 2002; Romanczuk et al., 2009; Vicsek et al., 1995), while our model explicitly implements the trade-off between random environmental and social cues. The previous simulation studies that considered similar mechanisms (e.g., when moving agents directly react to a non-conspecific cue as a predator (Demšar et al., 2015; Olson et al., 2016) or a food-patch (Couzin et al., 2005)) allowed the moving agents to access non-social information without a permanent cost, i.e. the trade-off only exists with a present environmental cue. In the proposed model, the vigilance limits the general capacity to react to social cues and thus affects the individual behavior even in the absence of environmental cues. Another major difference to most existing models (Calovi et al., 2018; Couzin et al., 2002; Jolles et al., 2017; Vicsek et al., 1995) is that the individual speed is not a model parameter but emerges from the interplay of burst rate, strength, duration and the tendency to follow environmental cues. Especially, accounting for the possibility of individual burst being triggered by either social or environmental cues adds a novel, yet ecologically relevant behavioral dimension to our agent-based model.

It is difficult to check if the predictions based on the simulated scenarios are valid. I already discussed some obvious limitations in Sect. 12.3, e.g. zebrafish as unfit model species for fishery, uncertainty about social forces in larger groups. In principle, the cohesion of fish shoals under large size selective mortality need to be measured and compared to regions with normal harvesting (using nets with small mesh sizes). If the resolution of sonar, used in commercial fishing, gives a fair estimate of the real shoal density is unclear. Even if the density could be estimated with sonar, cohesion could be influenced by site dependent by water temperature or salinity.

A maybe simpler indirect measurement of cohesion is the change in population of natural predators of the commercially fished species. A naive assumption is that high commercial fishing pressure

would reduce the prey abundance of natural predators, which in turn would reduce the population of the natural predator. However, if the predator population is increasing in regions of high commercial fishing compared to regions with less fishing pressure, it could be explained by a change in cohesion of the prey that reduces the confusion effect and is at the end beneficial for the natural predator. This could be done for example with the Galapagos sea lions (Zalophus wollebaeki) where a good population database exists (Riofrío-Lazo et al., 2017). However, it needs to be compared to the history of commercial fishing pressure, if fishing is size selective or not, and to other influences as the water temperature. It also needs to be known where the natural predator populations hunts. Thus, it is probably easier to perform more laboratory experiments to confirm our predictions.

The predictions are definitely relevant for recreational fishery, because large harvesting is actively performed to preserve the fish stock (Heino et al., 2015). However, it is questionable if it is relevant for commercial fishing with trawling nets. First, most regulations do not even support the use of nets with a large mesh size, i.e. all fish are caught irrespective of their size which renders commercial fishing as an unfit example of increased large-size selective mortality (Salomon and Holm-Müller, 2013). Second, if nets are used that reduce the by-catch and allow smaller fish to escape, it probably has an overall beneficial effect for the whole ecosystem and the fish stock. It preserves diversity and a part of the targeted swarm. Thus, the prediction results should be taken with caution and are not meant to undermine the efforts to make fishery more sustainable by reducing by-catch with large mesh sizes.

Part III

Quantifying collective escape waves of sulfur mollies

The thesis began with a theoretic model system in which a standardized predator attacks and individuals evolve according to their fitness, and continued with the analysis of experimental selection lines under increased size-dependent mortality and a model fit to those selection lines. Both parts analyzed different aspect of predation: (i) the direct evolutionary adaptation and (ii) the survival under predation if an evolutionary adaptation with respect to another context preceded. However, predator interactions are always modeled and the thesis lacks the reference to real systems. This gap will be closed by this part of the thesis.

It is in general difficult to monitor real predator prey-group interactions in the wild. The prey groups need to be found and the predator needs to be present at the same time and not be influenced/scared by the observers. Due to the dynamic nature of the behavior, one needs to follow the trajecory of the shoal and the predator. There have been great attempts to monitor predator prey group interaction in the wild (Axelsen et al., 2001; Gerlotto et al., 2006; Handegard et al., 2012; Pitcher, 1996; Price et al., 2013; Similä, 1997). However, most of the work is based on sonar recording. It only allows the analysis of a part of the group and even then, it is difficult or impossible to identify individuals of the densely packed prey.

In a particular sulfidic spring eco-system in Mexico, prey groups can be easily observed at the surface in two dimensional shoals, at constant location and under frequent attack by predators. In the last part of this thesis, I will introduce this system with its predators and the fascinating collective escape dives of the sulfur mollies (*P. sulphuraria*). The introduction highlights the adaptive value of the collective escape waves. I describe methods to quantify the escape waves on a macroscopic scale (tenth of meters). With these methods, I analyze if the location or the stimulus can affect the collective wave characteristics. To further understand the microscopic interactions that lead to the macroscopic phenomenon, I analyze video recording with a field of view that allows the detection of individuals (maximal one square meter). Thereby, we gain insights on how structural properties influence the speed of the collective wave (polarization, orientation, density). The part is closed with a discussion and summary.

In this field system, I try to answer the following main research questions:

- Does the fish group responds differently to experimental manipulation compared to natural predators?
- How does the information transfer depend on the collective state and structure?

The exploration and analysis of the system is a large collaboration spanning different laboratories. I will explicitly refer for any figure, table, software, data that was not created/analyzed solely by me. The contributions from others is summarized in Sect. 17.1.

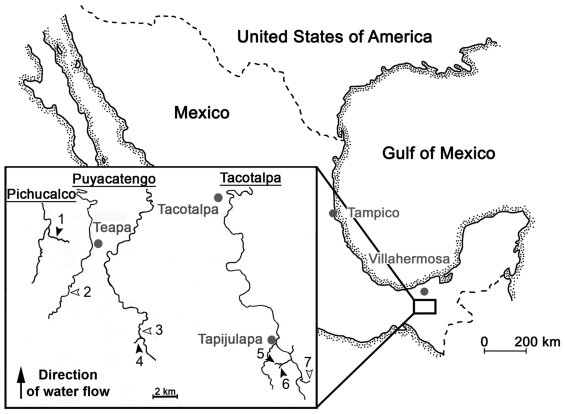

Figure 13.1: Map of endemic region of *Poecilia sulphuraria*. Sulfidic springs and non-sulfidic springs are marked with black and white arrows. The species *Poecilia sulphuraria* is endemic in 1 (Baños del Azufre) where all the data of this part of the thesis was collected. Figure adapted from Bierbach et al. (2018).

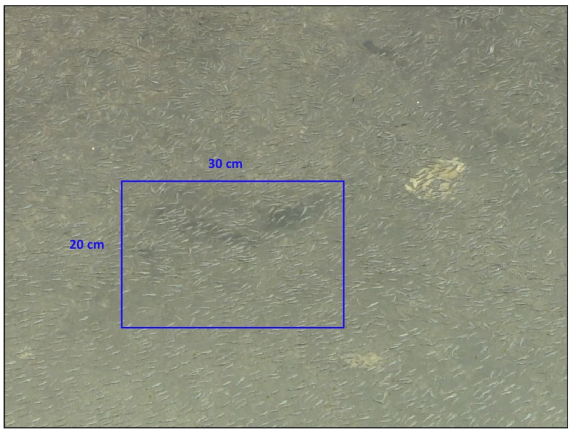

Figure 13.2: Density of *P. sulphuraria* in sulfidic spring, adapted from Doran et al. (2020).

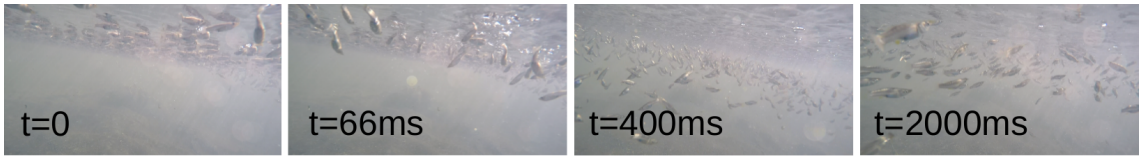

Figure 14.1: Snapshots of collective diving of *P. sulphuraria*.

14 Introduction

14.1 The collective escape dives of mollies in sulfidic springs

This part of my thesis is all around the fish species *Poecilia sulphuraria* that lives exclusively around a sulfidic spring, which is known as Baños del Azufre, close to the city Teapa in Tabasco, a state of Mexico (Fig. 13.1). I participated in three field trips (2017, 2018 and 2019) to collect data on the collective diving behavior of these fish.

The reactive nature of sulfur (H_2S) with oxygen creates a low oxygen concentration in the water. Additionally, H_2S binds to hemoglobin and therefore blocks binding sites of oxygen. This creates an extreme environment and only two fish species adaptad to the sulfidic water conditions at Baños del Azufre: *Poecilia sulphuraria* and *Gambusia eurystoma*. They can bind oxygen at much larger H_2S concentration than non-adapted species. However, the water has a very low oxygen concentration. It forces the fish to do aquatic surface respiration (ASR), i.e. they swim at the water surface where the water is directly in contact with the air and therefore has a higher oxygen content. Note, that the abundance of *P. sulphuraria* is much larger compared to *G. eurystoma* and initial observations suggest that *G. eurystoma* do not mix with *P. Sulphurria* and are located rather in different positions within the stream (Lukas, 2018). This leads to clear spatial separation of the species and in general to the system being dominated by large scale schools of *P. sulphuraria*.

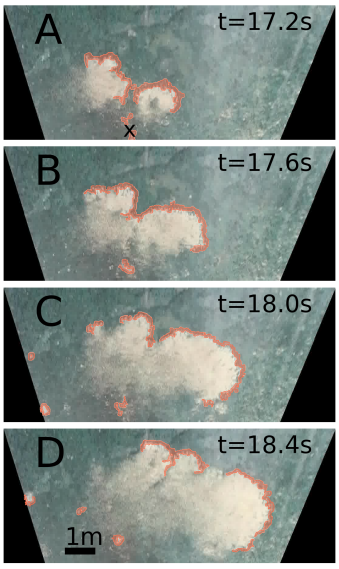

Figure 14.2: Snapshots of propagating collective startling wave with detected front (red).

The sulfidic springs are colonized by anaerobic bacteria, which form huge bacteria films on the river bed. These bacteria serve as food for the fish and its high abundance allows huge populations of *P. sulphuraria*. The fish are found in high densities at the water surface (Fig. 13.2) with up to 4600 individuals per square meter (Doran et al., 2020).

Since the extreme environment excludes any fish predators, only airborne predators are a threat. Due to the high abundance and reachability from outside the water (due to aquatic surface respiration) the fish shoals are often attacked by birds as the green kingfisher (*Chloroceryle americana*) and greater kiskadee (*Pitangus sulphuratus*), among others. They show a collective escape mechanism during which individuals startle from the surface to greater water depths and are copied by close conspecifics (Fig. 14.1). This behavioral mimicry can spread in a wave like manner for several meters involving thousands of fish (Fig. 14.2).

The system has some clear advantages: (i) Fish shoals are two dimensional carpets on the water surface, i.e. the whole shoal can be video recorded, which is difficult for three-dimensional shoals. (ii) The shoals can be observed at the same spot for a long period, i.e. no searching and/or following necessary. (iii) The shoal is under frequent attack by predators, i.e. a large number of

predator prey interactions happen in a short time interval, which provides large sample sizes.

14.2 The adaptive value of collective fish waves

The two most frequent bird predator are the green kingfisher (*Chlorocerule americana*) and the greater kiskadee (*Pitangus sulphuratus*), which have very different attack techniques. While the kiskadee only enters the water with its beak, the kingfisher dives fully in the water (Fig. 14.3). The kingfisher, as the common name suggests, specializes in fish as prey. Normally, fish swims in greater depth and its diving behavior during the attack on sulfur mollies, which swims at the surface (ASR), allows him to catch already escaping/diving fish.

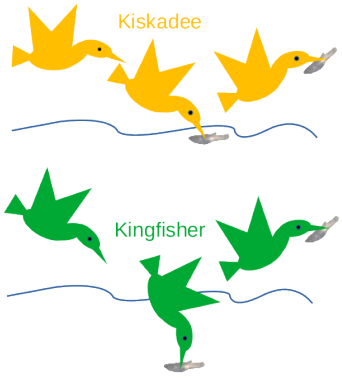

Figure 14.3: Illustration of kingfisher and kiskadee attack.

This strong impact creates water surface waves and an acoustic stimulus. It causes the shoal of fish to dive collectively and repeatedly, i.e. after the first collective diving wave the resurfaced fish perform another collective diving wave without an additional stimulus (attack). This collective escape behavior can be repeated up to 20 times (Fig. 14.4A). Thus, the system has strong similarities with traveling waves in excitable media (Meron, 1992).

In contrast, the kiskadee does not cause water surface waves and its beak produces much weaker acoustic signals compared to the kingfisher. Consequently, the sulfur mollies only engage in local diving escapes but not in global collective diving and no repetitions of it, as observed after Kingfisher attacks (Fig. 14.4A).

Doran et al. (2020) found that the collective, self-repeating fish waves have adaptive value for the fish. If fish repeatedly and collectively dive, the predators wait longer between subsequent attacks. This manifests in a longer waiting time between attack for the kingfisher compared to the kiskadee because only the former causes repeated collective escape waves (Fig. 14.4B). Importantly, Doran et al. (2020) showed that this effect is not species related by experimentally inducing self-repeating collective escape waves by shooting a small projectile with a slingshot into the fish shoal. Kiskadees also wait longer between attacks if confronted with these experimentally induced collective escape waves (Fig. 14.5), as previously observed for the kingfisher.

Despite the adaptive value of the behavior, it is unclear if fish have learned that repeated diving decreases the frequency of attacks or if it is an evolutionary stable behavior adaptation. Motivated by the strong resemblance of the system to excitable media, it is highly interesting to quantify this biological system of weakly related individuals and how it has reached this analogous state. It is difficult to keep the fish captive and even more difficult to keep them in large densities in a constrained environment. In fact, even constraining them in their natural environments, by enclosing shoals with fine nets, caused an extreme physical stress. Thus, up to now, field studies have been the only way to gather data about this fascinating system.

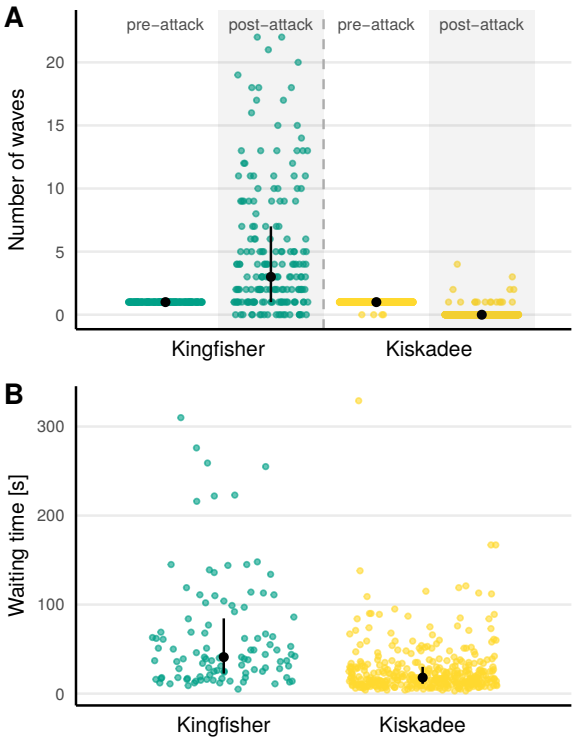

Figure 14.4: The number of pre- and post-attack waves (**A**) and waiting times between attacks (**B**) of kiskadee and kingfisher. Figure adapted from Doran et al. (2020).

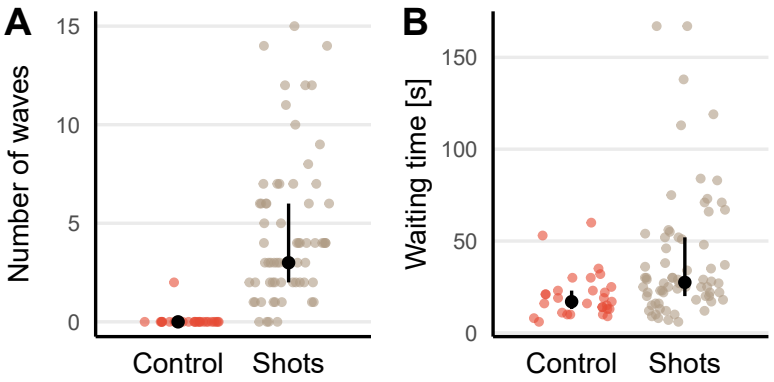

Figure 14.5: Number of experimentally triggered waves and control waves (**A**) and related kiskadee waiting time for both cases (**B**). The control corresponds to a projectile that hit the water surface outside of the fish shoal. It takes account for possible movement of the person shooting the projectile, which could scare the bird and therefore artificially prolong its waiting time. Figure adapted from Doran et al. (2020).

15 Quantifying the escape wave on the macro-scale

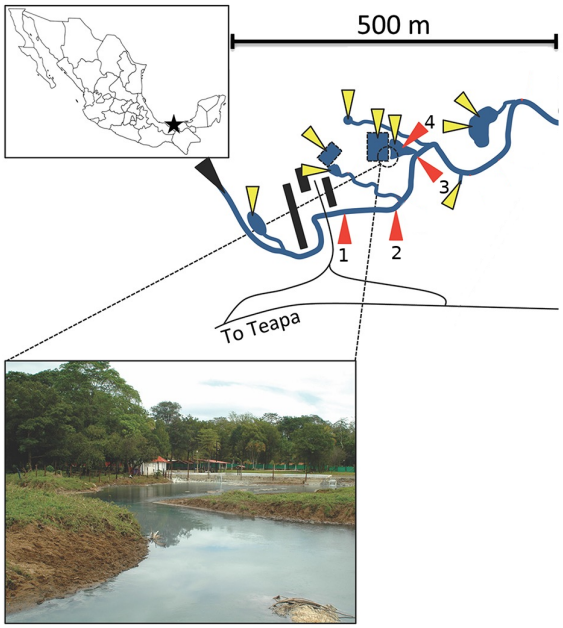

Figure 15.1: Selected sampling site in Baños del Azufre where video recordings where taken (red triangles). At sites 1-3 the water stream velocity is much larger compared to location 4, which we also refer to as *big pool* location. Yellow triangles mark springs with H_2S-rich water. Figure adapted from Culumber et al. (2016)

In order to know what influences the macroscopic behavior of the collective escape waves, they need first to be characterized. For this purpose, I introduce algorithms that detect the wavefront and the wave speed. We recorded at different locations, shown in Fig. 15.1, the collective response to different stimuli. With the quantification method I will investigate if the location or the nature of the stimulus effect the escape wave.

The shoal is stimulated either by a bird attack or a projectile shot with a slingshot in or close to the shoal. The recordings were taken by a CANON XF200 camera at 50fps full HD resolution or by a Sony HD Handycam from the shore.

15.1 Estimation methods for wave size and speed

For each analyzed video exists a calibration video in which a square-shaped calibration board (size of 1.55 x 1.55m) floats on the surface. This allows for a transformation (homography) to a top-view perspective and provides us with a conversion factor from pixel to meter, which is 200 pixels per meter. The final video is transformed to gray scale and cut to exclude regions that were too far away from the calibration board and therefore not correctly transformed.

The wave detection by its active area and the wave speed estimation was done in a close collaboration with Pawel Romanczuk and Carsten Rösner.

15.1.1 Wave detection

The waves are detected by (i) computing from the $M \times M$ area surrounding each pixel the baseline standard deviation, i.e. the baseline activity, (ii) comparing the current activity to the baseline activity and declaring it as active if it crosses a threshold T and (iii) applying a standard blob detection algorithm to the resulting 3D activity map. The latter assigns to all active pixels that are continuously connected the same wave label.

To determine the baseline activity the part of the video is used that was prior to a bird attack or an artificial stimulation. The baseline activity is a positive matrix of the same dimension as the spatial extend of the video $S_b \in \mathbb{R}^2_+$.

The activity map in frame t is a binary map defined by

$$A(t) = (S(t) - S_b) > T \tag{15.1}$$

with $S(t)$ as the activity map in frame t and T as the mentioned threshold.

Since the recordings are from different days and day times, the light conditions changed. To guarantee a detection of the continuous wave fronts (Fig. 14.2), different parameter combinations of M and T were used: $[M, T] = [11px, 10],\ [12px, 4.5], [10px, 6]$.

Of course, this activity detection mechanism will also detect any floating objects on the water surface, as bacteria films, which loosens constantly from the riverbed. To distinguish floating objects from waves, I excluded all spatial blobs (computed by a standard blob detection mechanism) with an area below $400cm^2$. The remaining blobs are merged by the same blob detection mechanism, except that it is now applied to the entire video instead of single frames. At the end, small waves with an unique area covered below $1m^2$ are excluded. These are local waves caused by small perturbations as bubbles.

15.1.2 Wave-speed estimation

To estimate the wave speed of each detected wave, we computed the contours of the wave in the current and past frame. From the pixels that define the contour of the current frame and that are not within the contours of the past frame (Fig. 15.2), the shortest distance to the any contour of the past frame is computed. The absolute value of this distance vector $\vec{v_i}(t)$ for pixel i in frame t is averaged over all distance vectors in this frame to obtain the average wave speed of the corresponding wave. The average wave velocity is then averaged over all frames in which the wave exists.

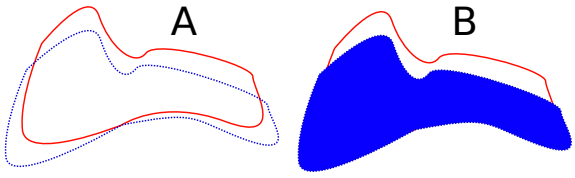

Figure 15.2: Illustration of wave speed computation. **A:** contours of the present (solid red) and past (dashed blue) activity map. **B:** non-overlapping present contour (solid red) from which each pixel is compared to the past contour.

The estimation of the wave speed via the shortest path between pixels of the current to the last contour of the wave front is based on the assumption that the wavefront follows the Huygens-Fresnel principle, i.e. that each point of the wavefront is the source of a spherical (circular in the 2D case) wavelet.

Note, that this methods assumes that the frame rate of the video is high enough to allow the detected wavefront of the current and past frame to overlap in space. This allows to define the relevant wave front by considering automatically the traveling direction. If the frame rate would be so low that both contours do not overlap anymore, the wave speed would be underestimated.

15.2 Wave-speed dependence on shoal location and stimulus

15.2.1 Stimulus variation

Lukas et al. (2020) showed that for *P. sulphuraria* the strength of the response depends on the nature of the cue. They found that the fast startle diving speed and duration vary between a visual, acoustic and combined cue. The attack by a kingfisher corresponds to a multimodal cue. To be able to study the system in detail and not to rely on kingfishers to perturb it and initiate collective escape waves, we shot bio-degradable small object in the water (dried peas or skittles shot via a slingshot or an analog construction). Note that there are a large number of kingfisher attacks. However, the Kinfisher attacks at its own will and not necessarily in the camera's field of view. It is not possible to change the camera view because for each new camera angle a new calibration is necessary for which the system is disturbed heavily (place calibration board on water surface). To

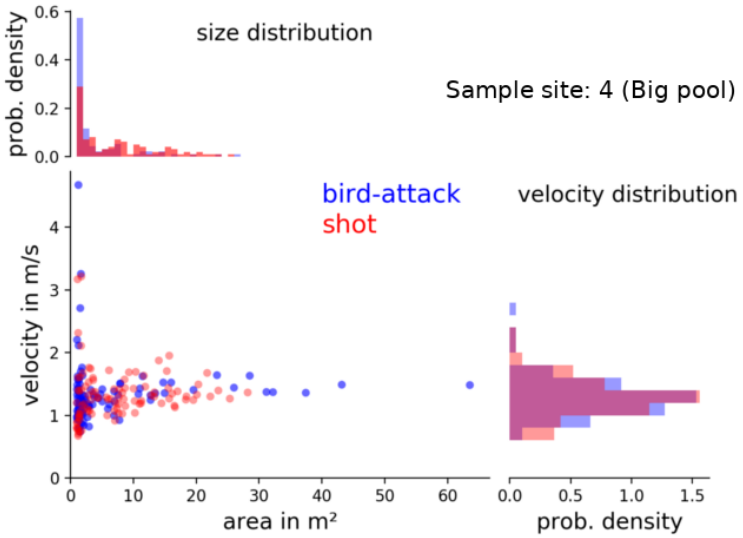

Figure 15.3: Area and speed of bird (blue) and shot (red) initiated waves (location: big pool).

ensure that the artificial stimulation mimics the real kingfisher attack, we compared the area of the waves and its propagation speed between the rare kingfisher attacks and the numerous artificial stimulations with shots.

In Fig. 15.3 the unique are covered by the wave and the wave speed show no apparent difference in between the two stimuli. The mean wave speed initiated by a kingfisher attack is $\langle v \rangle_{kf} = (1.33 \pm 0.49) m/s$, which is quantitative equivalent to the wave speed initiated by an artificial shot, $\langle v \rangle_s = (1.32 \pm 0.40) m/s$. The maximal average wave velocity is $max(v)_{kf} = 4.67 m/s$ for kingfisher initiated waves and $max(v)_s = 3.21 m/s$ for shot initiated waves.

The average unique wave area initiated by a kingfisher is $\langle A \rangle_{kf} = (6.2 \pm 10.1) m^2$ and by a shot $\langle A \rangle_s = (7.8 \pm 6.7) m^2$. Note that the maximal wave area by bird attacks is much larger than by wave initiated attacks. This is because the kingfisher attacked often at the periphery of the camera's field of view. Thus, I was forced to analyze a larger field of view for the kingfisher attacks, which was not the case for the artificial shots (we always aimed at the center). Therefore, larger wave areas could be detected for waves stimulated by the kingfisher.

15.2.2 Variation in location

It is important to know, how the sampling site influences the wave characteristics. If this were the case, we could filter out the specific element of a site that affects the wave. Some selected sites are highlighted in Fig. 15.1. Especially site 4, which I will refer to as *big pool*, is different from the sites 1-3. The riverbed of the big pools is much larger and therefore the water stream velocity lower compared to the other sites. In addition, the frequency of strong external perturbations was lower at the big pool. A lower attack/perturbation frequency prevents the response (collective diving) of one perturbation to overlap with the response to another cue.

If the sites 1-3 (river) are compared with waves observed at site 4 (big pool) the distribution of the unique area covered by the waves is similar. The waves initiated by birds are smaller compared to shot initiated waves (Fig. 15.4A,B). The difference can be explained by the larger area of the big pool compared to the river sites.

The sites have a much larger discrepancy in the velocity distribution with an average speed of $\langle v \rangle \approx 0.7 m/s$ for the bird attack initiated waves in the river compared to the speed in the big pool $\langle v \rangle \approx 1.3 m/s$, which is larger by about a factor of 2 (Fig. 15.4B,b). Interestingly, the difference almost vanishes if the wave speed is only estimated from velocities parallel to the shore

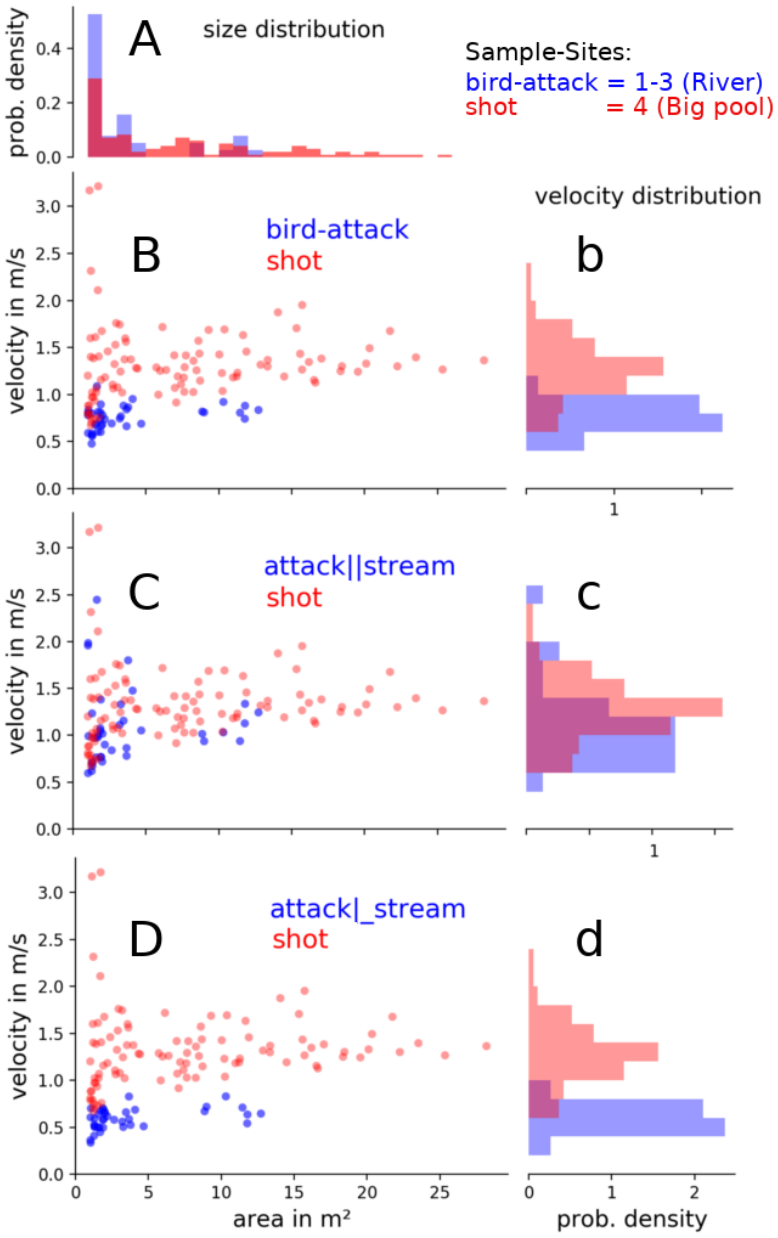

Figure 15.4: Site comparison in area and speed of bird and shot initiated waves. Bird attack initiated waves (blue) were recorded at sites 1-3 (river) and shot initiated waves (red) were recorded at site 4 (big pool). **A:** Distribution of unique area covered by waves. Each wave is characterized by its unique area covered and its average wave speed. The speed of waves initiated by bird attacks is estimated from all velocities (**B,b**), only from the velocity parallel to the shore/stream (**C,c**) or perpendicular to it (**D,d**).

(Fig. 15.4C,c). The opposite holds for wave speeds estimated only from velocities perpendicular to the shore (Fig. 15.4D,d).

A possible explanation is that the collective escape dive wave propagates faster in the heading direction of the school. The river has a faster flowing velocity and the dense schools tend to keep a stationary position with respect to the shore. In order to keep their position they are heading upstream. There are two consequences: (i) the school is strongly polarized compared to the big pool site and (ii) the aspect ratio of the fish (around four) causes a different nearest neighbor distance parallel and perpendicular to the polarization. Thus, the nearest neighbor to the front is larger than to the side and, in consequence, the speed parallel to the shore is faster.

15.3 Summary and discussion

The introduced algorithm allowed a quantification of the collective escape waves of the sulfur mollies on the macro-scale. The average speed and it's unique area covered were computed. The maximal average wave speed was $\langle v \rangle_{kf} = 4.67 m/s$, i.e. this wave had peak velocities even greater than this value. The maximal startle velocity of the related species *Poecilia reticulata* is $v_{startle} \approx 1.2 m/s$ (Oufiero and Garland, 2009). Thus, the collective wave is by far faster than the swimming speed of an individual fish, also if not the maximal but the average wave speed is considered.

We did not find differences in wave speed or area between the natural stimulus (kingfisher attack) and an artificial stimulation (slingshot). This is to be expected, because collective escape wave reach individuals which have not sensed the initiation stimulus directly. Thus, in principle the stimulus can not alter the characteristics of the social wave. However, this is only true if the startling behavior is binary. This is likely the case for the startle response, because the behavior is triggered by the Mauthner cell, a command neuron, and is highly stereotypic (Nissanov et al., 1990).

The finding that experimental manipulation triggers a collective response that is undistinguishable from responses to natural stimuli, makes the system even more valuable.

We could explain the site dependend speed of the collective waves by differences in the water stream velocity of the sites. The higher stream velocities at the river sites force the fish to polarize with a group heading parallel to the shore. This structural element slows down wave with a traveling direction perpendicular to the shore.

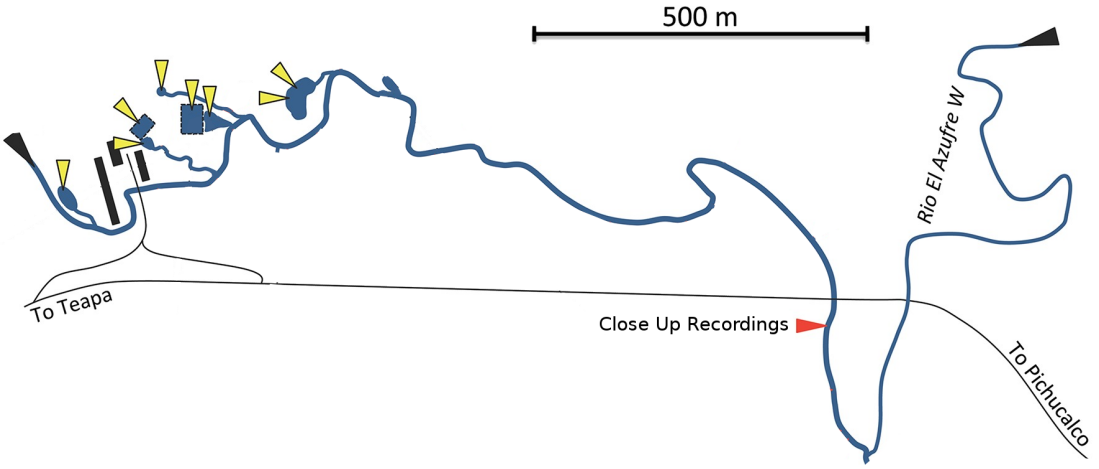

Figure 16.1: Location of close up recordings marked by red arrow. Figure adapted from Culumber et al. (2016).

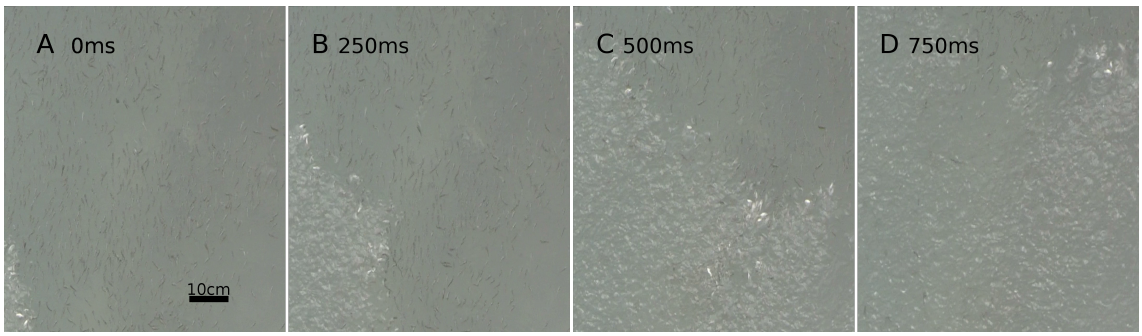

Figure 16.2: Collective escape wave with detectable individuals. The collective escape wave starts in the lower left corner (**A**) and needs 750 milliseconds to reach the upper left corner (**B-D**). Individual fish are detectable in the single frames.

16 Collective escape waves on the micro/meso-scale

In the previous chapter, video recordings were analyzed with a field of view in the order of meters. No individuals were detectable but the propagation of the wave was monitored and analyzed on the macro-scale. Here I focus on the analysis of video recordings with a field of view as large as possible to still recognize individual fish. A typical example is shown in Fig. 16.2. We recorded the shoal (i) without manipulation and (ii) with manipulation by shooting a small bio-degradable projectile close but outside of the cameras field of view. The recordings were taken roughly 1 kilometer downstream, as indicated in Fig. 16.1.

To capture the behavior at the micro-scale, the smallest units need to be monitored, i.e. individual fish must be detectable. However, since the escape of a single fish disturbs the water surface and therefore obscures the detection of its neighbors, we are forced to analyze the meso-scale. In the presented approach, the meso-scale corresponds to an average value over a small area. Thereby can an area unit contain none, only one (micro-scale) or multiple individuals, depending on the

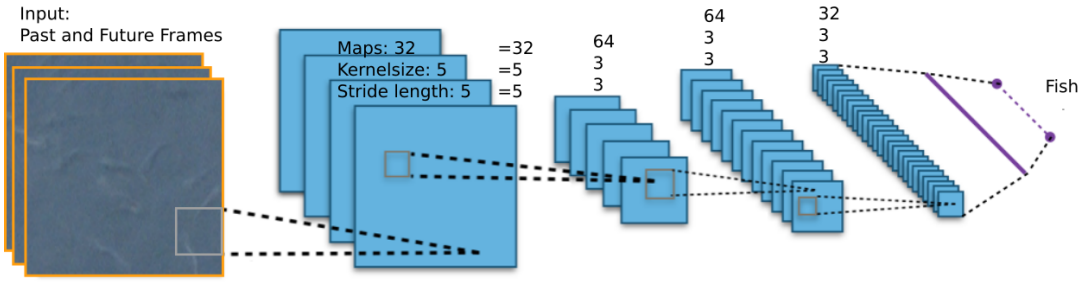

Figure 16.3: Architecture sketch of Convolutional Neural Network to detect fish. To predict if a fish is in the center of a batch, four convolutional layers are applied. The input of the network corresponds to a smaller section of the original image and some past and future frames of the same section. In the first convolutional layer 32 feature maps a created by convolving a kernel of size 5 with the batch of each frame. The stride length, which is the number pixel the kernel is shifted each step, always corresponds to the kernel size. Note that only the convolutions are shown and not any others steps (normalization, up sampling, etc.). The number of feature maps, the kernel size and the stride length is depicted above each layer.

local density. Of course, it would be desirable to remain on the micro-scale for a better mechanistic understanding of the process (the meso-scale could be additionally investigated in means of, for example, network properties) but video recordings of the escape wave from outside the river restrict the analysis to the meso-scale.

The existing animal tracking tools (e.g. Francisco et al., 2020; Graving et al., 2019; Pérez-Escudero et al., 2014) are not suitable for tracking the video recordings to the extremely noisy and variable field video recordings with weak contrast, frequent light reflections, permanent perturbation of the water surface, which makes individual fish hard to detect even for a trained observer.

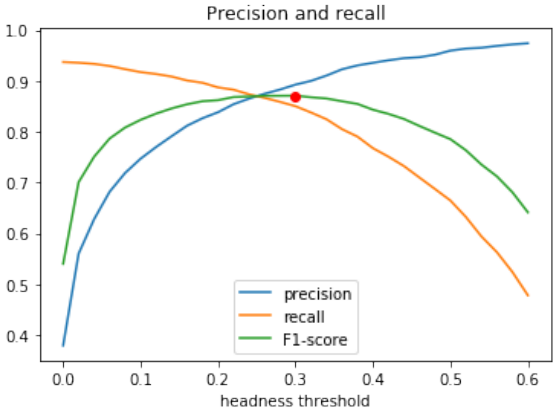

Figure 16.4: Precision recall and F1-score of trained CNN.

In the following, I will introduce different, partly extending methods to access the micro- and meso-scale. A convolutional neural network (CNN) is introduced in order to detect individual fish with a F-score of about 0.88. An algorithm, based on the output of the CNN, links the independently detected fish positions to trajectories and a length criterion for the trajectories can push the F-score to around 0.92. For the wave detection on the meso-scale, the already presented macro-scale wave-detection mechanism is extended. It assigns for each coarse grained unit a wave speed and direction that agrees best with the model of a local planar wave. Finally, I analyze how local structural properties influence the propagation of the collective wave.

16.1 Methods

16.1.1 FishNet: individual fish detection via a convolutional neural network

FishNet is a convolutional neural network (CNN) developed by Leon Sixt from the Landgraf-group and is only applied in this thesis. Therefore, I will only sketch its architecture and performance.

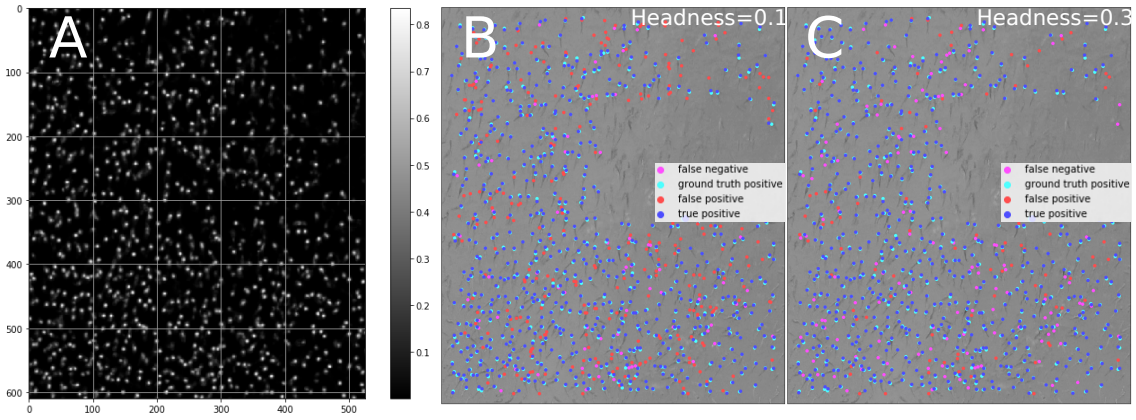

Figure 16.5: From the headness values (**A**) discrete predictions are generated by applying a headness threshold. For a prediction with a headness above the threshold the location is classified to represent a fish. Comparison of predictions with ground truth, visualized by true and false positive and negative predictions with the ground truth for a threshold of 0.1 (**B**) and 0.3 (**C**).

The CNN is illustrated in Fig. 16.3 and consists of four convolutional layers of which each has between 32 and 64 feature maps. The input of the CNN are batches/sections of the original frame of fixed size (side length of 128 pixels) and should be large enough to contain always a whole fish. In order to predict if a fish is in the center of a batch, the network also uses the past and future frames of the same section as input. Another network is trained to predict the direction where the fish is pointing to, its orientation. The architecture of the orientation network is identical to the one that predicts if a fish is in the batch center (Fig. 16.3).

We generated training data by annotating images with the position and orientation of fish. After 150 training epochs, the loss saturated and does not deviate strongly from the loss of the validating data set (Fig. C.1). The final network generates headness values (Fig. 16.5A) which are converted into fish predictions by a threshold criteria, i.e. if the headness is above a threshold the position is classified as a fish location (Fig. 16.5B,C). The precision ($p = TP/(TP + FP)$) increases and the recall ($r = TP/(TP + FN)$) decreases with increasing the headness threshold (Fig. 16.4). We defined the optimal threshold value via the maximum of the F1-score, the geometric mean of precision and recall $F1 = 2/(r^{-1} + p^{-1})$, which is $F1 = 0.87$ at a headness threshold of about $t_h = 0.3$.

16.1.2 FishLink: from fish detection to trajectories

The previous section introduced *fishNet*, which predicts for each frame of a video the location of fish. Together with Conor MacBride (DAAD-RISE intern) we developed a linking algorithm that extends *fishNet*. It links the detected fish across frames, i.e. it creates trajectories. The algorithm also allows fish to disappear for a certain number of frames f_d, these missing frames are then interpolated. Fish might vanish for a few frames due to prediction errors of *fishNet* and/or due to the noisy data (water surface waves create reflections or worsen the contrast).

The algorithm first connects fish in neighboring frames that have the highest probability to belong together. The probability that the fish i_f in frame f and fish j_{f+1} of the next frame represent the same fish p_{same} is the value of a two dimensional Gaussian distribution

$$p_{same}(i_f, j_{j+1}) \propto \exp\left(-\frac{d(i_f, j_{f+1})^2}{\sigma_d^2} - \frac{\alpha(i_f, j_{f+1})^2}{\sigma_\alpha^2}\right) . \tag{16.1}$$

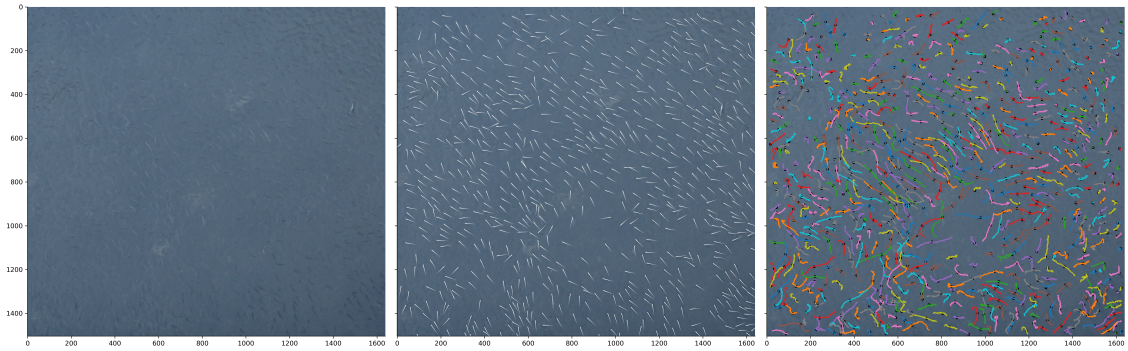

Figure 16.6: *FishLink* example of trajectories in a frame with length threshold $l_t = 17$. The first image is the original image, the second shows predicted fish and their orientations and the last frame shows their trajectories.

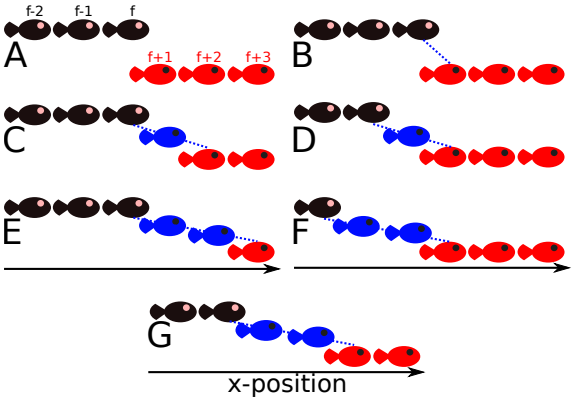

Figure 16.7: **A**: two trajectories are shown where one ends at frame f (black fish) and the other starts at frame $f+1$ (red fish). **B-G**: possible links between both trajectories are displayed. It is assumed that in maximal f_d frames the trajectories could be corrupted. Thus, it is allowed to connect pairs across maximal f_d frames (**C-G**). The probability that both trajectories represent the same fish is the largest probability p_{same} among all pairs be the same fish. If the largest probability is a pair that assumes corrupted frames, the position of the fish in the respective frames is linearly interpolated (blue fish).

With the distance between the two detections $d(i_f, j_{f+1}) = |\vec{r}(i_f), \vec{r}(j_{f+1})|$ and their orientation difference $\alpha(i_f, j_{f+1}) = \angle(\hat{v}(i_f), \hat{v}(j_{f+1}))$ as the dimensions. The standard deviation σ_d, σ_α for both dimensions needs to be set as an additional parameter to the model.

Note that the terminology for the orientation as the normalized velocity $\hat{v} = \vec{v}/v$ is not exact, since the water stream velocity causes the heading direction and velocity direction to differ. At the next step, the algorithm computes for each ending trajectory at frame f and each trajectory that starts in the interval $[f-f_d+1, f+f_d]$ the probability that they represent the same fish. For each pair of ending $E(f)$ and starting trajectory $S(f_s)$ this probability p_{link} corresponds to the highest probability of all possible combinations of the two trajectories assuming that maximal f_d frames can be missing or be false positives. An example of all possible combinations of two trajectories (one ending at f and the other starting at $f+1$) is depicted in Fig. 16.7.

If all possible trajectory combinations are listed, the most likely is established and then the second most likely (excluding all combinations between already combined trajectories) and so forth. Finally, to increase precision and decrease the recall, as for *fishNet*, a threshold length l_t needs to be set. If the trajectory is longer than l_t, it is considered to represent the movement of a real fish. An example output for a frame predicted by *fishNet* and from which the trajectories are recreated by *fishLink* is shown in Fig. 16.6.

FishLink has therefore 4 additional parameters (f_d, σ_d, σ_α, l_t) and since it is based on *fishNet* the total number of parameters is five (headness threshold). The headness threshold should be chosen below the optimum value. This allows the algorithm to keep true positives and sort false positives out based on the length criterion. With a headness threshold of 0.2 the F1-score is

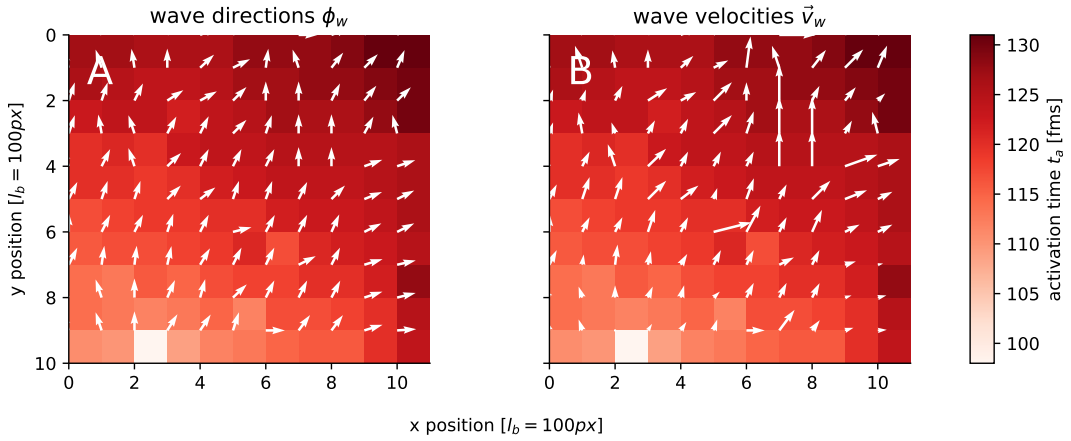

Figure 16.9: Activation time t_a and velocity map of collective escape wave. The activation time for each box is color coded. The corresponding wave direction (**A**) and speed (**B**) is estimated by assuming local planar waves and indicated for each box by a white arrow.

maximized at a length threshold of about $l_t = 17$ (Fig. 16.8). Note that the maximal F1-score of *fishLink* is by 5 percent greater than the maximum in *fishNet*.

In summary, *fishLink* increases the F1-score and generates trajectory data from position data. However, this comes with a larger computation cost (the linkage of 100 frames takes 3 minutes with parallelized computation on 40 cores) and 4 extra parameter, which needs to be set correctly. The performance of *fishLink* is expected to decrease if the following 2 conditions hold: (i) the frame-rate is too low with respect to the velocity of the fish and (ii) the fish are in a disordered state. The linkage method is based on closest position in consecutive frames and if fish are disordered, their trajectories can cross. A too coarse time resolution will result in a false linkage of these crossing trajectories.

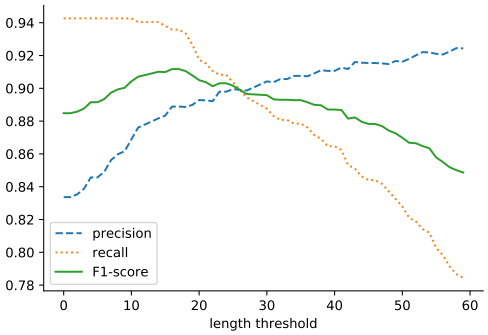

Figure 16.8: Precision, recall and F1-score of *fishLink* using a headness threshold of 0.2.

16.1.3 Estimating local wave speed and direction

When a collective escape wave is passing through the field of view, it is not possible to define the exact time an individual starts to participate in the wave (Fig. 16.2). This is due to the startle response of individuals that disturb the water surface. If an individual startles, it is not possible to detect if it's neighbors startle as well or do so delayed. It forces us to analyze the system on the meso-scale. The cameras field of view is partitioned in boxes of length $l_b = 100px = 6.6cm$. The box size represents a compromise between the number of data points (spatial resolution) and the resolution of the wave-speed. The smaller the box size the larger the number of coarse grained data points. However, the fixed frame rate ($fps = 50$) in combination with smaller box lengths, decreases the speed of the fastest detectable wave. From the macro-scale analysis (Sect. 15.2) we know that the average wave speed is $\langle v_w \rangle \approx 1.5m/s$. The box length $l_b = 6.6cm$ allows a sufficient fine detection in this region (for more details see Sect. C.1).

We create for every box an activation time t_a which is when the collective escape wave reached this box. To estimate the location of the collective escape wave front we use the wave detection

method described in Sect. 15.1.1. We adapted the size of the window M that surrounds each pixel to $M = 32px = 2.1cm$. This corresponds to typical length of *P. sulphuraria* and should not be set smaller. Otherwise, the movement of a fish could trigger an activation of the focal pixel. The resulting activation map is resized to one value per box. The activation time t_a corresponds to the first time the box is activated.

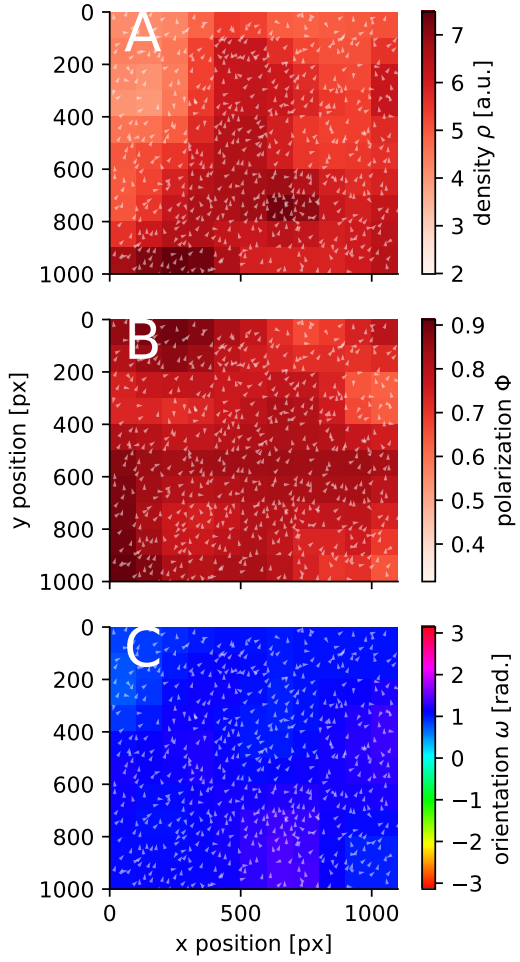

Figure 16.10: Meso-scale averages of density (**A**), polarization (**B**) and orientation (**C**) of a fish shoal. The white arrows mark individual fish from which those measures are computed.

We approximate the local activation pattern of the neighborhood of each box (the focal and its eight neighbor cells) with a planar wave with the velocity vector $\vec{v}_w = v_w[\cos\phi_w, \sin\phi_w]$. For each possible wave velocity vector $\vec{v}_w$ exists a theoretical prediction for the activation time $t`_a(j)$ of the neighboring box $j \in \mathbb{N}_i$ if the wave passes the focal box i at its measured activation time

$$t`_a(j, \vec{v}_w) = \frac{(\vec{r}_j - \vec{r}_i) \cdot \hat{v}_w}{v_w} \tag{16.2}$$

$$= \frac{(\vec{r}_j - \vec{r}_i) \cdot \vec{v}_w}{v_w^2} \ . \tag{16.3}$$

Note that in the first line $\hat{v}_w$ is the normalized wave velocity and therefore the nominator represents the distance of the neighboring box to the focal box projected onto the wave direction $\hat{v}_w$. The specific wave velocity of the focal box is the one that minimizes the error between the predicted and measured activation time

$$error(i) = \sum_{j \in \mathbb{N}_i} |t_a(j) - t`_a(j, \vec{v}_w)| \ . \tag{16.4}$$

For the optimization, the Nelder-Mead method is applied. Note that neighbors are excluded from the error if they are not activated at all. If the focal box was never activated, it is excluded from the whole analysis.

A typical activation time map with the optimized wave velocities is shown in Fig. 16.9. Note that the wave direction is pointing in the majority of the cases in the directions $\pm[0, \pi/4, \pi/2, 3/4\pi, \pi]$. Directions in between are rare and the favored directions correspond exactly to the eight relative neighbor positions. This suggests that the optimizer tends to minimize the sum of errors in Eq. 16.4 by reducing the two summands that represent a specific neighbor pair at opposite relative positions (e.g. front-and-back neighbors, left-and-right neighbors, upper-left-and-lower-right neighbors). Probably, the optimizer selects the neighbor pair that is in best agreement with the planar wave assumption. Nonetheless, the wave direction and speed is approximated to a fair degree by this method.

16.2 Results: collective escape wave dependence on density and orientation

The introduced methods allow us to compare how structural properties of the fish shoal, obtained from the individual fish positions and orientations (*fishNet*), influence the wave speed. Since the wave-speed analysis returns coarse grained data in boxes of size $l_b = 100px = 6.6cm$ (Sec. 16.1.3),

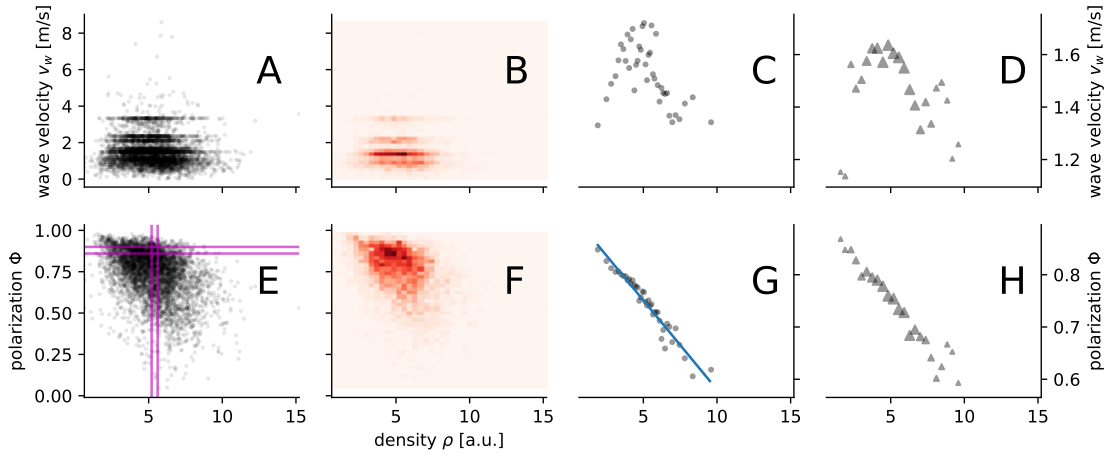

Figure 16.11: Dependence of wave velocity v_w (**A-D**) and polarization Φ (**E-H**) on density ρ. **A, E**: raw data points. **B, F**: 2D histogram of the raw data. **C, G**: accumulated data with each data points representing the same number of raw data points (equal weights). **D, H**: binned data in equidistant bins with the size of the triangles representing its weight. The blue line (**G**) is a linear regression on the data. The raw data between the two vertical and horizontal magenta lines in **E** is used to disentangle the influence of ρ and Φ on v_w, respectively.

the structural properties of the shoal are also coarse grained to the same scale. This is done by computing the polarization for the box i with its center at $\vec{r}_i$ to

$$\Phi(i) = \left| \frac{\sum_j w(r_{ij}, \sigma_c)\hat{v}_j}{\sum_j w(r_{ij}, \sigma_c)} \right| \tag{16.5}$$

with $\hat{v}_j$ as the heading vector of a fish j, r_{ij} as the distance between fish j and box center i and $w(r_{ij}, \sigma_c)$ as the weight. The weight is a Gaussian weight with standard deviation $\sigma_c = l_b = 100px$:

$$w(r, \sigma_c) = e^{-r^2/(2\sigma_c^2)} \ . \tag{16.6}$$

The orientation angle $\omega(i)$ for box i is computed analogously. For the density $\rho(i)$ the weights stay the same but the normalization is based on a weighted area, represented by a grid with grid positions $\vec{r}_g$

$$\rho(i) = \frac{\sum_j w(r_{ij}, \sigma_c)}{\sum_g w(r_{ig}, \sigma_c)} \ . \tag{16.7}$$

An illustration of the coarse grained measures is shown in Fig. 16.10.

We analyzed in total 55 collective escape wave events, which adds up to roughly 5 636 coarse grained measures. The comparison of the wave velocity v_w with density ρ reveals an ideal density at which the wave propagates fastest (Fig. 16.11A-C). This is in agreement with an underlying interaction network based on the Voronoi tessellation with an additional decrease of link weight with increased distance. If the escape wave can only spread from one fish to its Voronoi neighbors, it will slow down in regions of high densities, as observed. However, if the distance between fish is too large (low densities), they are connected only weakly and the information transmission is hampered (Rosenthal et al., 2015). These two features cause the speed to peak at intermediate densities. Note that the discrete horizontal lines in the raw data (Fig. 16.11A,B) are caused by the already mentioned peculiarity of the wave velocity estimation (Sect. 16.1.3). Please find a more detailed explanation in Sect. C.2.

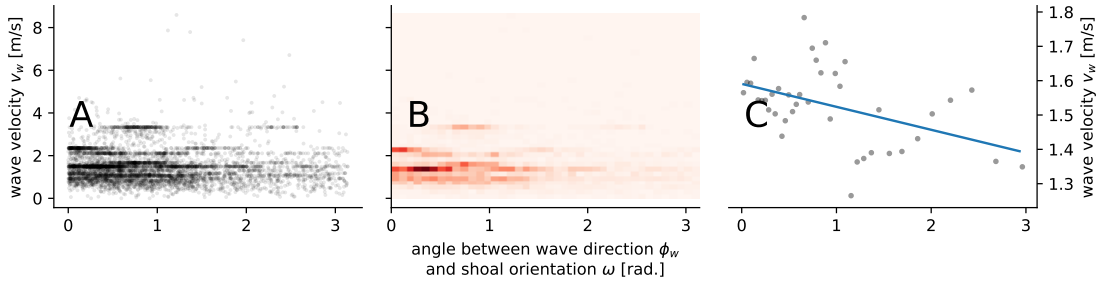

Figure 16.12: Wave velocity v_w dependence on angle between shoal orientation ω and wave direction ϕ_w. **A**: raw data. **B**: 2D histogram of the data. **C**: accumulated data with each data points representing the same number of raw data points (equal weights) and a linear fit (blue line).

The polarization Φ depends as well on the density ρ; however, in a rather unexpected way, i.e. the polarization linearly decreases with increasing density (Fig. 16.11D-F). The correlation coefficient (R-value) is $C(\rho, \Phi) \approx -0.29$ and the regression analysis suggests a p-value of zero. With an increased density one would expect that the fish is forced to be stronger aligned with each other because of their large aspect ratio of about 5. This means, fish can be packed closer without colliding if they are more polarized. However, the distribution of the polarization suggests that the fish is already strongly aligned in most boxes. Thus, packing might not be a problem in this coordination regime. The reason for the negative correlation is rather the Gaussian kernel used to estimate the local polarization (Eq. 16.5). At low densities, the sum in Eq. 16.5 is dominated by the fish with the heaviest weight. Since the polarization of a single individual takes by definition the maximal value one, the polarization increases with decreasing density. This dependence makes it unclear if the wave speed v_w depends on the density ρ or on the polarization Φ.

To clarify how the density ρ and the polarization Φ influence the wave speed I repeated the analysis with two subsets of the data. I split the data in 10 equally large parts sorted with respect to the polarization Φ. With the part that has the smallest standard deviation in polarization (data between two horizontal magenta lines in Fig. 16.11D) the dependence analysis of v_w on ρ is repeated (Fig. C.3). The same procedure is applied on the data subset with the smallest variation in density ρ (data between two vertical magenta lines in Fig. 16.11D) as shown in Fig. C.4. It turns out that both measures influences the wave speed (compare panel C with F of Fig. C.4, and panel D with H of Fig. C.3). It means the wave speed is positive correlated with the polarization (linear regression: $r = 0.21$, $p < 10^{-6}$) and has an optimal density at which the wave speed is largest.

We found in the Sect. 15.2 that the wave speed measured on the macro-scale was higher for a propagation direction parallel to the shore. We hypothesized that the collective escape waves propagates faster in the direction parallel to the shoal orientation. To validate this hypothesis I compute for each box the angle between the wave direction ϕ_w and the shoal orientation ω. The wave speed shows a clear decrease with an increase in angle discrepancy (Fig. 16.12). A linear regression analysis supports this finding with $r = -0.07$ and $p < 10^{-5}$.

17 Summary and discussion

The collective escape waves of *P. sulphuraria* were analyzed on the macro and micro-scale. On the macro-scale we tracked the propagation of waves that span and travel meters and involve tenth of thousands of fish. We developed algorithms to detect wavefronts and to estimate their speed and area. It turned out that an artificially triggered wave is indistinguishable from a wave triggered by a kingfisher attack. The location of the shoal can influence the wave speed if the water stream velocity differs between locations. It turned out that the difference in speed is caused by a slower propagation perpendicular to the shoals heading direction. To understand the microscopic mechanism of the wave propagation we analyzed recordings with a small field of view that allowed the detection of individuals. The individual fish detection was automated by a convolutional neural network *fishNet* for which I introduced an extension algorithm *fishLink* that links the detections of single frames to trajectories and allows missing links. From the individual detections the local group structure was computed (density, polarization, orientation). The local wave speed and direction was estimated by the velocity vector of a planar wave that bests corresponds to the observed wave propagation on a local scale. Finally, I found that the wave speed depends non-linear on the density and linear on the polarization. I could support the macroscopic finding that the wave is fastest if its propagation direction is aligned with the local shoal heading.

The finding that the collective shoal responds to an artificial stimulation in the same manner as to a natural bird attack allowed us to justify our experimental manipulations (Doran et al., 2020).

Of specific importance are the mechanistic relations between shoal structure and collective information propagation. The fact that an optimal density exists where the wave spreads fastest is in line with the claim that realistic modeling of social interactions is not possible with purely metric models. In a metric model, where the response only depends on the distance and no occlusion exists, the wave speed should increase with density. A Voronoi network, which is in good agreement with a visual network (Strandburg-Peshkin et al., 2013), could explain the peak well.

The wave speed increases with polarization. It suggests that the fish with a higher social attention in one context (normal swimming behavior) follow social cues in a different behavioral context (fast startle response) faster.

I found that the wave speed depends on the angle difference between the wave direction and the local shoal orientation, which could even be observed on the macro-scale. It means that a wave propagates faster if it approaches individuals from behind. Additionally, most waves did approach the fish from behind (Fig. 16.12B). Initially, I hypothesized that the aspect ratio of the fish in a polarized swarm causes the nearest neighbor distance in the heading direction to be larger and therefore increases the wave speed. However, if this would be true, the wave speed would increase again for angle differences (between wave direction and shoal orientation) larger than $\pi/2$. Since this is not observed (Fig. 16.12C), the explanation is likely to be wrong.

An alternative explanation is that fish can not access information from the back as good as information from the front and therefore are more responsive to any social information from this direction. If this would be true, fish would have a complex decision making structure in which they weight information differently depending on the source of the information and their own personally accessible information. In this way, the signal of the conspecifics extend the sensory range of an individual.

The chosen meso-scale approach has clear limitation but was the most obvious choice for recordings from the water surface. The detection of fish in underwater recordings is not affected by the startling of conspecifics and should be considered as an alternative. However, underwater recordings pose new problems (the occlusion of diving fish) and from the technical side it is more elaborate. If the shoal moves out of the field of view of the underwater camera, the whole system needs to be perturbed because a new set up involves someone entering the water.

Another clear limitation is the local approximation of the escape wave as a planar wave. The

resulting raw data reveal patterns, which are artifact of the discrete nature in space and time of the wave velocity approximation on the course grained meso-scale representation of the system. This is likely the largest weak point and the results discussed here should be taken with caution. Until another approach has not verified the findings, they should be considered as preliminary. An alternative could have no fixed box-size but an adaptable size according to the speed of the wave-front at this position. The wave front could be estimated analogously to Sect. 15.1.2. Fast wave speeds would only locally coarsen data. In regions of low wave speeds, the spatial resolution would be increased.

It might be very insightful to model a contagion process on the boxes similar to the model by Dodds and Watts (2004). This approach was already successfully applied to model contagion processes of escape startles in fish by Sosna et al. (2019). Following their analysis, a cleaner approach would be to model the contagion process not on the coarse grained scale but directly on the network of individuals. The output could be coarse grained to compare the model to the experimental findings. If we stay on the meso-scale, the spread of the collective escape wave could also be modeled via cellular automata. In this case, each box represents one cell and the probability to be activated depends on the activity of the neighbors. However, one would lose information if the density is discarded. Therefore, the state of each cell is not only defined by its activity but also by the number of fish represented by this box. An additional dimension could be the longest delay to respond to a neighboring cell. This would encode the strength of the signal and would be in line with findings by Lukas et al. (2020) who showed that the escape response depends on the stimulus. With rules/probabilities to be activated we could pose questions like "Is the delay of a neighboring cell inducing a delay in the focal cell?". If we could answer this question with "yes", it could explain why repeated collective escape waves diminish over time.

The system of *P. sulphuraria* who perform ASR and escape via a fast startle response in the depth is extremely useful to study the propagation of information without density waves. In other systems, the escape trajectory is directed most often into the shoal. This creates density waves (Herbert-Read et al., 2015; Sosna et al., 2019). However, in this system fish flee without approaching/colliding with conspecifics. Therefore, it is an unique system that clearly deserves more attention. Our findings allow some very interesting hypothesis. Especially that fish weight social information stronger if it represents information to which individual has restricted access to, is exciting.

17.1 Authors contributions

The macro-scale video recordings were recorded at the river sites (1-3) by Juliane Lukas, Carolina Doran, Marie Habedank and Jens Krause. At the big pool site (site 4) Pawel Romanczuk, Haider Klenz and me were involved. The wave detection algorithm (Sect. 15.1.1) was mainly written by Pawel Romanczuk. It was extended to enable the processing of larger waves by Carsten Rösner and me. The wave speed estimation for the macro-scale recordings (Sect. 15.1.2) was developed by Carsten Rösner, Pawel Romanczuk and me. The macro-scale recordings were processed and analyzed by me (Sect. 15.2).

The *fishLink* algorithm (Sect. 16.1.2) was developed by Conor MacBride and me. The local wave speed estimation (Sect. 16.1.3) by a planar wave was developed by Jannik Thuemmel, Noam Miller and me. Most macro-scale data was preprocessed by Noam Miller and the final analysis was done by me.

Part IV

Conclusion and outlook

In this work, I have investigated different systems in which individuals move in a coordinated way. Not only the collective aspect of self-driven agents/individuals is a connecting element between the three parts of my work. Also, the biological question of how individuals or the group behaves in case of danger is a core overarching element. However, the different parts cover different approaches and methods to gain fundamental insights on collective predation response.

In the evolving collective (1st part) only models and simulations were used. It opened the possibility to test a hypothesis that would be difficult or impossible to handle with experimental methods. I could work out that the critical point under predation pressure and individual evolution is not only unstable but also accelerates evolution. The theoretical framework allowed me to identify its reason as the maximum spatial self-sorting of individuals according to their relative behavioral parameters at the phase transition. This could hardly have been achieved with experimental methods. The obvious drawback is that the obtained statements are only valid as long as the assumptions of the model are correct and complete.

The zebrafish model (2nd part), however, had a direct relation to experimental data. Numerical modeling of a mechanism could show that the otherwise unintuitive behavioral differences between the selection lines are consistent with the assumption that individuals can not react to social and environmental information simultaneously. Furthermore, the model representation of the selection lines allowed me to predict how each line would survive in a natural predator and industrial fishery context.

However, for the collective escape waves of the sulfur fish (3rd part) no generative model was developed. The reason for this is mainly that already the quantification of these waves was a great challenge, which inspired intense discussions and provided various opportunities for cooperation. A corresponding systematic modeling study of the system goes beyond the scope of this thesis, but is currently being performed by other lab members. The escape waves are generated by fishes of a few centimeters in size but spread over several meters. The quantification of the waves at the macro and micro level has complemented each other. We observed on both scales that the local orientation of the swarm influences the speed of the wave. It suggests that fish is more responsive to social information to which it personally has only limited access to.

In the following, I revisit aspects of the three parts and highlight commonalities and use their differences to discuss the results with a broader view.

The concept of evolution is fundamental for the first two parts and allows an interesting outlook in relation to the third part the work. In the first part, it turns out that evolution under predation pressure leads to strongly ordered collectives. The model representations of the zebrafish lines were based on groups that were subject to a very different selection pressure, namely size-dependent mortality. Although this selection is motivated by human intervention (recreational fishing, fisheries), it can also be associated with natural predators (e.g. bass preferred smaller stoneroller; Theodorakis, 1989). However, the evolution under human selection pressure is extremely interesting and relevant, since overfishing of natural stocks is the standard rather than the exception. An evolutionary study based on a spatially explicit fishing process does not exist but would also

be linked to a number of assumptions. For example, Guerra et al. (2020) assume that industrial fishery targets large shoals and therefore reduces the spatially explicit fishing process to a conceptual one. But group size says nothing about the structure, i.e. two groups with the same number of individuals could occupy a different volume (density differences). Less cohesive shoals could be harder to catch, because larger nets would be necessary. This could create an evolutionary pressure which favors individuals with a lower social cohesion. As a consequence, these less cohesive individuals could be at higher risk under natural predation (disable the confusion effect).

The last study raises the question of which selection pressure has evolutionarily produced the collective escape dives. In the context of the first study, which tested the criticality hypothesis, it would be extremely relevant to know how close the collective behavior is to a critical point. If the system is near to such a point, it could complement the first study. A comparison of the two could lead to minimal conditions necessary to enable self-organization to criticality. These minimal conditions would be of interdisciplinary interest (physics, neuroscience, gene regulatory networks, evolutionary biology). A possibility for such a minimal condition, which is very relevant for group living, weakly related animals, is multi-level selection (within- and between-group selection, Wilson, 1975, 1997).

The other concept that plays a fundamental role in all studies is the structure of the collective and how it is related to the behavior rules, the phase transition, the transfer of information. In the first part, the structure is very multifaceted. On the one hand, the particular dynamics at the phase transition cause a minimal density: correlated velocity fluctuations lead to homogeneous subgroups that strive in different directions, thus pulling the swarm apart. On the other hand, the symmetry breaking in combination with a high mobility of the individuals in the collective leads to a maximum self-sorting. The former leads to the collective having the greatest distance to the hunter at the phase transition, a passive effect. Also in the second study, the structure is a passive advantage or disadvantage compared to a natural hunter. Only in the last study, it is shown that the structure is extremely important for the information transfer. It should be noted that the structural properties by itself (density and polarization) affect the general transfer of information. However, also the structure relative to the information is important, e.g. the polarization with respect to the propagation direction of the information has an additional influence on the information transfer.

The latter is associated with the assumption that individuals weight social information more if they have a restricted sensing to the direction from where the information comes. Also in the second study, the experimental differences between the selection lines could be explained by assuming that individuals weight the available information differently. The weighting was a direct consequence of the cognitive limit of not being able to process all information (social and environmental) simultaneously. Therefore, it highlights the importance of information processing at the level of individuals and the role of potential cognitive and sensory constraints.

Retrospective, I would like to emphasize how enriching the interdisciplinary character of this work was. It allowed me to connect concepts from different fields. It becomes clear which elements allow the transfer of a concept and which make it necessary to expand or even discard it. Thus, the already known can be completely rediscovered and this on both sides. On one side, a system can be analyzed with a new concept/tool/method. On the other side, one gets to know the already known concept anew and questions the fundamental assumptions.

Appendix

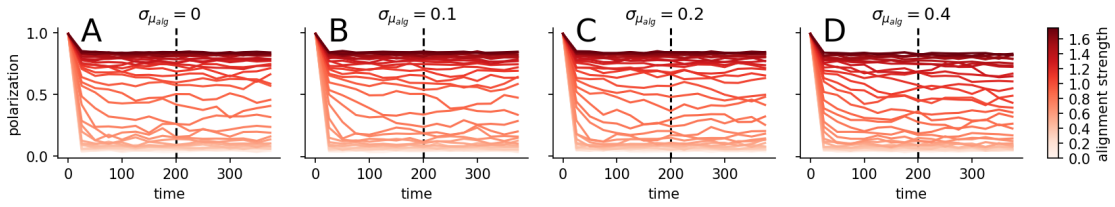

Figure A.1: Equilibration in polarization for different standard deviation of the population. The panels (**A, B, C, D**) correspond to $\sigma_{\mu_{alg}} = [0, 0.1, 0.2, 0.4]$ respectively. The line color codes the alignment strength and varies between zero and $\langle \mu_{alg} \rangle = 1.6$. The vertical dashed line indicates the equilibration time $T_{eq} = 200$, i.e. the time after which the measurements are taken.

A Evolution in a collective under predatory pressure

A.1 Equilibration time for the prey collective

The prey-collective is initiated in a slightly perturbed ordered state and equilibrates for $T_{eq} = 200$ time units. Only after the equilibration any measurements are taken and the predator appears. This equilibration time is sufficient for the collective to relax as can be seen in the polarization (Fig.A.1A). Note that the relaxation is slower the closer the system is to the order-disorder transition, i.e. at medium polarization around $\Phi = 0.5$. This critical slowing down is typical for dynamical systems close to a bifurcation. The equilibration time is still sufficient if heterogeneity is introduced by varying the standard deviation of the population(Fig.A.1B-D).

A.2 Numeric stability

To ensure the results are not compromised by artefacts caused by numeric imprecision this section deals with the algorithmic stability using the Euler-Maruyama method to simulate a stochastic differential equation. The time-step dt should be much smaller than the persistence time $\tau_p = 2$, smaller than the shortest correlation time and small enough to fulfill the stability criterion and to avoid oscillating behavior.

A general linear stochastic differential equation (here in Langevin form)

$$\frac{dx}{dt} = \mu x + \sigma \eta(t) \tag{A.1}$$

which is simulated via the Euler-Mayurama method

$$x_{n+1} = (1 + \mu dt) x_n + \sigma \sqrt{dt} \eta(t) \tag{A.2}$$

is algorithmic stable Gardiner (2004, p.379) if

$$|1 + \mu dt| < 1. \tag{A.3}$$

The above stability criterion is simply a Lyapunov-stability for discrete deterministic processes. It is in the same time the criterion for errors, between the real process and the simulated process, diverging with time. A more strict criterion is that the process approaches its steady state continuously from above or below, i.e. it is not allowed to alternate/oscillate between the sides

$$0 < 1 + \mu dt < 1. \tag{A.4}$$

An even stricter criterion is that the time step should be much smaller than the correlation time of a process active in the system, i.e. restricting the time step to be smaller than a tenth of the correlation time

$$\frac{1}{10|\mu|} \leq dt. \tag{A.5}$$

Note, that this ensures that the process is not just uncorrelated Gaussian white noise (see A.3). Also note this inequality does not try to "prevent" the low force limit, where the noise term dominates the agent's velocity, because I consider here the correlation time of stable states. For example, consider an agent strongly aligned with its neighbors (clearly far from the low force limit) whose velocity is perturbed by the noise term. The Eq. A.5 prevents the agent to just fluctuate randomly around the group direction of motion, i.e. it allows the agent to relax towards the group direction according to its correlation time.

The above concepts are applicable to the discretization of a linear SDE. To apply the stability criteria to the high-dimensional non-linear stochastic process (Eq.3.1) I consider the maximal angular change of a focal agent by an isolated social-force. The consideration of isolated forces is reasonable because the distance force is pointing in general in a different direction as the alignment force, e.g. for strongly aligned agents the alignment force is parallel to its current velocity but the distance force is pointing in the opposite direction as the closest neighbor. Without loss of generality I rotate the system such that the considered force is always pointing in the x-direction. I also assume that the force is stable.

The distance regulating- $\vec{F}_d$ and and flee-force $\vec{F}_f$ depend on the spatial position of the agent and therefore only implicit depend on the heading direction. Thus, the dependence on the heading direction is introduced only via the projection of these forces on the direction of angular change $\hat{e}_{\varphi,i} = [-\sin\varphi_i, \cos\varphi_i]$ which results in

$$\frac{d\varphi_i}{dt} = \mu_{x,i} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} -\sin\varphi_i \\ \cos\varphi_i \end{pmatrix} + \sqrt{2D_r}\eta(t) \tag{A.6a}$$

$$\frac{d\varphi_i}{dt} = \mu_{x,i} \sin\varphi_i + \sqrt{2D_r}\eta(t). \tag{A.6b}$$

The above equation is non-linear but with respect to stability/convergence I can substitute the sine-function with a linear dependence

$$\frac{d\varphi_i}{dt} = \mu_{x,i}\varphi_i \ . \tag{A.7}$$

This results in stronger forces, shorter correlation time and a stricter condition. Note that $x \in \{\mu_d, \mu_{flee}\}$.

The alignment force $F_{alg,i}$ depends explicitly on the heading-direction and since large forces correspond to an ordered state I assume that (i) the neighbors of the focal agent are strongly aligned with each other and (ii) the deviation in heading direction of the focal agent is small. Thus,

$$\frac{d\varphi_i}{dt} = (\mu_d \sum_{j\in\mathbb{N}_i} \vec{v}_j - \vec{v}_i) \cdot \hat{e}_{\varphi,i} + \sqrt{2D_r}\eta(t) \tag{A.8a}$$

$$= \mu_d(\langle\vec{v}\rangle_{\mathbb{N}_i} - \vec{v}_i)) \cdot \hat{e}_{\varphi,i} + \sqrt{2D_r}\eta(t) \tag{A.8b}$$

$$= \left(\mu_d \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} \cos\varphi_i \\ \sin\varphi_i \end{pmatrix}\right) \cdot \begin{pmatrix} -\sin\varphi_i \\ \cos\varphi_i \end{pmatrix} + \sqrt{2D_r}\eta(t) \tag{A.8c}$$

$$= -\mu_d \sin\varphi + \sqrt{2D_r}\eta(t) \tag{A.8d}$$

$$\approx -\mu_d\varphi + \sqrt{2D_r}\eta(t) \tag{A.8e}$$

where I used in the last step the second assumption, i.e. deviations from the group-heading direction are small.

Thus, every force on a focal prey can be approximated by a linear SDE for which the stability consideration eq.A.5 needs to hold.

A.3 Numerical stability estimate

In Sect. A.2 I introduced the inequality Eq. A.12 which ensures that the process can relax in its steady state in a fashion that preserves its correlation time. It therefore prevents that the process to resemble white noise. I will illustrate this with a simple approximation.

At each time-step a Gaussian-white noise term is added with a variance of $dt\sigma^2$. For too large dt the noise term could dominate the dynamics and the stochastic process would resemble an uncorrelated Gaussian white noise. First I estimate the maximum value away from the stationary state to which the process can in principle be stochastically driven x_{max}. At this value the deterministic reduction $dt|\mu|x_{max}$ equals a standard noise term

$$dt|\mu|x_{max} = \sigma\sqrt{dt} \tag{A.9}$$

$$x_{max} = \frac{\sigma}{\sqrt{dt}|\mu|}. \tag{A.10}$$

Now I ensure that this maximal value is always larger than the standard noise term

$$x_{max} > \sigma\sqrt{dt} \tag{A.11a}$$

$$\frac{\sigma}{\sqrt{dt}|\mu|} > \sigma\sqrt{dt} \tag{A.11b}$$

$$\frac{1}{\sqrt{dt}|\mu|} > \sqrt{dt} \tag{A.11c}$$

$$\frac{1}{|\mu|} > dt\ . \tag{A.11d}$$

The above equation states that the time-step should be smaller than the correlation time of the deterministic part. Note that the inequality Eq. A.11d is equivalent to Eq. A.4 and therefore not conservative at all. Also is it unlikely that the stochastic process reaches regularly its theoretic maximum x_{max}, therefore I take a more strict condition by restricting the time step to be smaller than a tenth of the correlation time

$$\frac{1}{10|\mu|} \leq dt\ . \tag{A.12}$$

A.4 Detailed robustness analysis for the group response

In the main text the main findings with respect to the group measures are: (i) susceptibility peaks at the transition, (ii) the distance to the predator as well but (iii) the maximum vanishes if a response measure is used that controls against the spatial structure (e.g. escape ratio R_{esc}). Here I investigate how these results change if parameters of the prey-model or the predator model are changed. The robustness analysis with respect to the flee strength was already done in detail in the main text. Note that the flee strength is in this respect a special parameter because it represents the connection from the predator to the prey shoal, i.e. the signal strength.

A.4.1 Prey-parameter dependence

In the precedent section only the alignment strength and the angular noise have been changed. These parameters have been investigated closely because they trigger the order disorder transition. However, as summarized in Tab. 3.1, these parameters are only 2 of the 4 prey-model parameter (speed v_0 and favored distance r_d are used for the dimensionless and are not counted). The count does not include the flee strength μ_{flee} because it is strictly speaking a predator-prey parameter. The two remaining "pure" prey parameter are the force distance regulating strength μ_d and the its slope m_d with which the distance regulating force switches from attraction to repulsion force. Both measures, μ_d and m_d, regulate therefore the strength with which the agents try to stay at

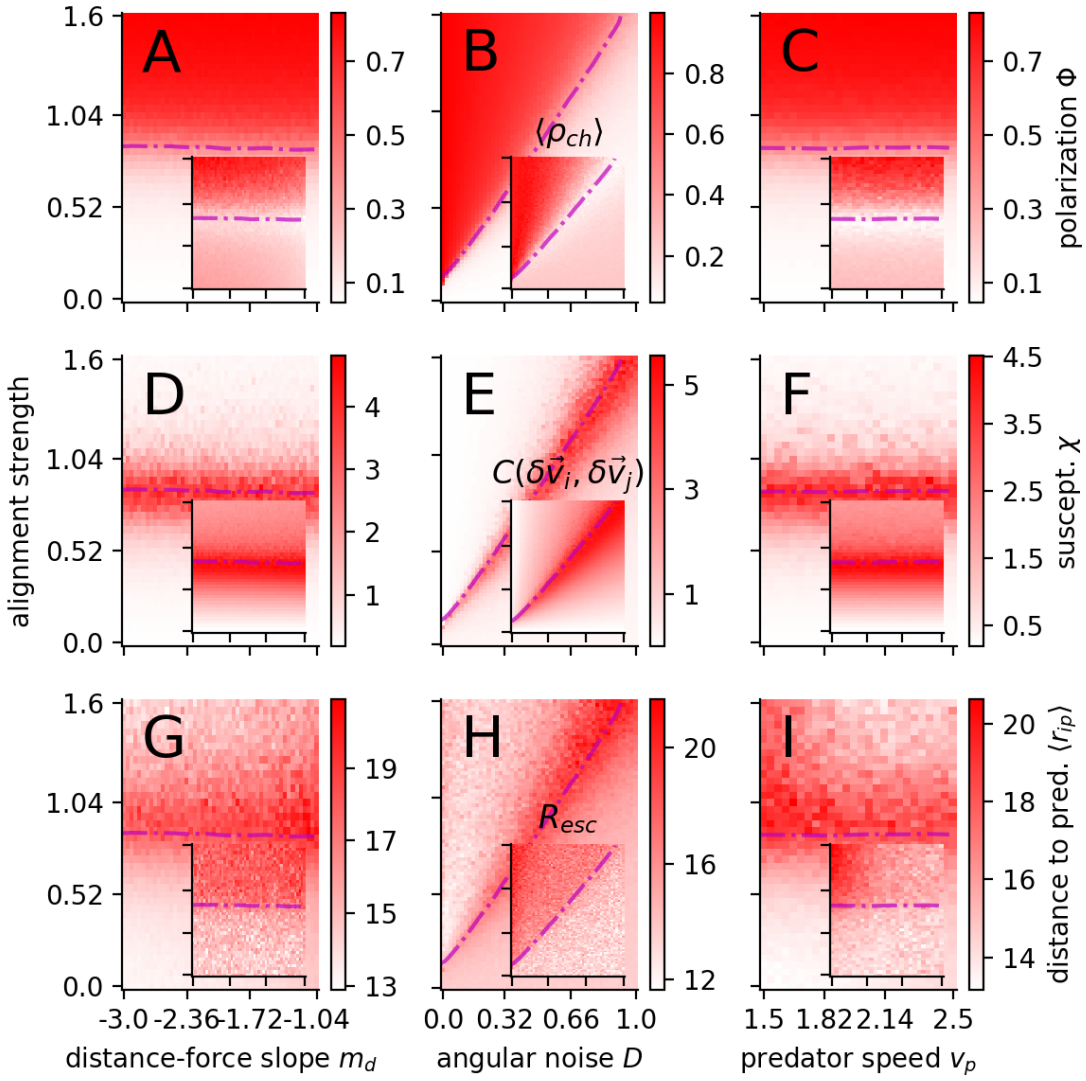

Figure A.2: Influence of predator speed and steepness of distance regulating force on Group measures. **A-C**: polarization Φ and the density estimated by the area of convex Hull $\langle \rho_C H \rangle$ as inset in each panel. **D-F**: susceptibility χ and the local information transfer estimated velocity fluctuation correlation between interacting agents $C(\delta\vec{v}_i, \delta\vec{v}_j)$ as inset in each panel. **G-I**: average distance to the predator $\langle r_{ip} \rangle$ and the escape ratio R_{esc} as inset in each panel. The slope of the distance regulating force m_d (1st column), the angular noise D (2nd column) and the predator speed v_p (3rd column) have been varied with respect to the default parameters listed in Tab. 3.1. The parameters listed in Tab. 3.1 for m_d and v_p correspond to the center values of the respective x-axis.

distance r_d to each other and can roughly be considered as analog. Variation in μ_d does not change the reported results as shown in Fig. A.2A,D,G. Therefore, the results can be considered robust against changes in prey-parameters.

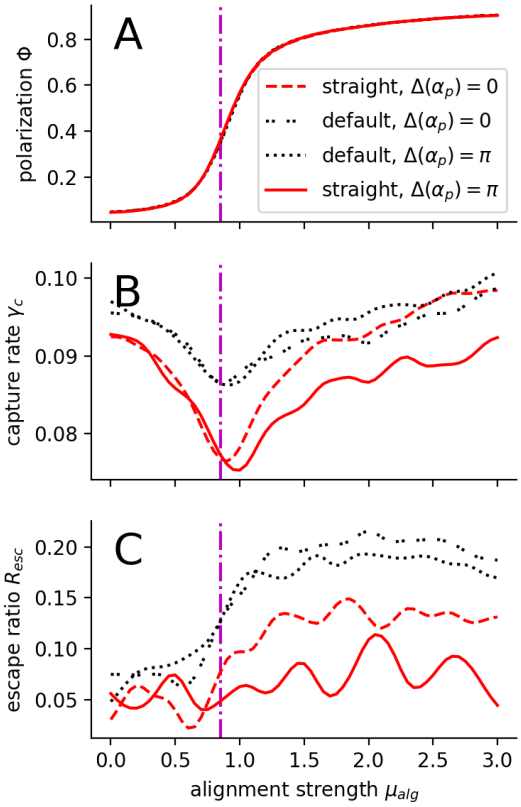

Figure A.3: Group response under alternative predator attacks. The polarization Φ (**A**), capture rate γ_c (**B**) and the escape ratio R_{esc} (**C**) for a default (black line) and straight (red line) attacking predator. The attack also varies in the standard deviation of the initiation angle of the predator $\Delta(\alpha_p)$.

A.4.2 Predator variations

The predator was chosen to be twice as fast as the prey. There are obvious consequences for a much faster or slower predator. If the predator is slower than the prey, it will never catch any. If the predator is faster, the prey has no chance to respond to it, e.g. a decrease in distance to the predator is expected. In Fig. A.2I the distance does decrease with increasing predator speed v_p and as does the escape ratio R_{esc}. All other group measures unrelated to the predator are, of course, unaffected (Fig. A.2C,F).

To see, if the predator dynamics and the resulting trajectory have an effect on the results, I implemented a predator that follows the center of mass of the shoal until it is close and then keeps its direction. For this predator the simulation is finished if it passes the shoal. The predator passed the shoal if no prey is in front of it. I will refer to this predator as the "straight" predator.

Additionally, the angle relative to the group velocity at which the predator is initiated α_p is varied. In the default attack the angle is random, i.e. the predator appears randomly in front, behind or at the side of the shoal. I test this case against a predator that always appears behind the shoal (indicated by $\Delta(\alpha_p) = 0$ in Fig. A.3).

These predator attack variations do not influence qualitatively the finding that the group response (e.g. the capture rate Fig. A.3B), is best at the transition, but if I control against the structure, the response increases with the order parameter (e.g. the escape ratio R_{esc}, Fig. A.3C). The randomness of the attack angle has little effect on γ_c and R_{esc}. However, the "straight" predator has a consistently lower capture rate than the default predator (Fig. A.3C), which is

reasonable because the default predator follows the closer prey and therefore should increase its capture. Note that an estimation of the response via the distance to the predator produces equivalent results (Fig. A.4). The only difference is that for a specific predator attack (straight without angular variation) a wall effect can be detected, i.e. the fleeing prey build up a dense wall in front of the predator (see Sect. A.4.3).

Thus, also changes of the movement, and initiation behavior of the predator does not affect the results. Interestingly, the escape ratio R_{esc} only increasing until an alignment strength of $\mu_{alg} \approx 2$ and than slowly decays with increasing order (Fig. A.3C), which is also the case for the structure controlled distance (Fig. A.4C). This has not been reported before, because this analysis considers larger alignment strength than before. It suggests that there is a responsive group optima, next to the structural group optima at the order disorder transition. This will be clarified in the next section where I increase the signal strength and analyze even larger alignment strength.

A.4.3 Straight attacking predator and the wall-effect

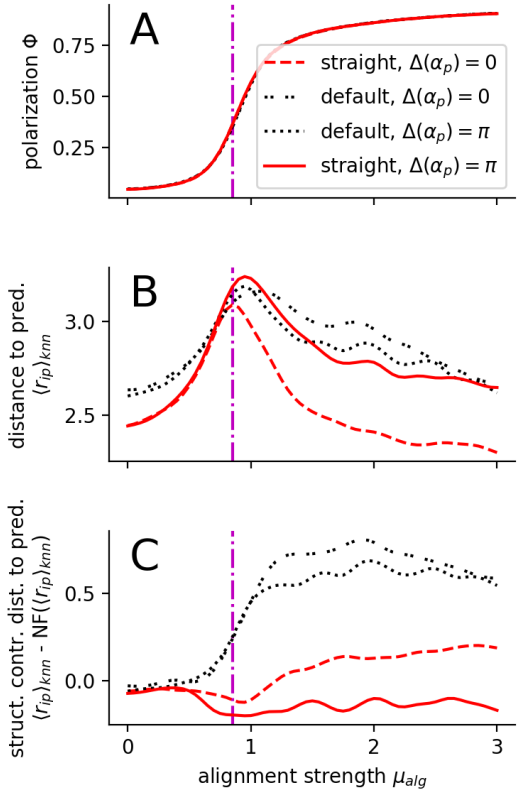

Figure A.4: Group response under alternative predator attacks (distance measures). The polarization Φ (**A**), distance to the predator averaged over its frontal k-nearest neighbors $\langle r_{ip} \rangle_{knn}$ with $k = 10$ (**B**) and structure controlled version of this distance computed by subtracting the distance of non-fleeing (NF) prey (**C**) for a default (black line) and straight (red line) attacking predator. The attack also varies in the standard deviation of the initiation angle of the predator $\Delta(\alpha_p)$. For a zero standard deviation the predator attacks always from behind.

In Sect. A.4.2 the group measures are discussed and how they change if the predator attack scheme varies. The discussion is based on the capture rate and the escape ratio. Here I show in Fig. A.4 that the same holds for distance related measures as the distance to the predator

averaged over the frontal k-nearest neighbors $\langle r_{ip}\rangle_{knn}$. There is one important difference that is not visible for the capture rate and escape ratio: The structure controlled distance to the predator averaged over the frontal k-nearest neighbors $\langle r_{ip}\rangle_{knn} - \mathrm{NF}(\langle r_{ip}\rangle_{knn})$ takes negative values for the straight attacking predator. That means fleeing agents are closer to the predator than their non-fleeing analogs. This seems counterintuitive; however, the escape ratio is not negative in the same region (Fig. A.3C). This means the fleeing prey escape better but remain closer, which can be well explained by the wall-effect. The fleeing prey initially increase their distance but can not accelerate and therefore form a denser wall in front of the predator. Since the measure is topological, the higher density lowers the mean distance of the fleeing agents.

A.5 ESS: balancing social vs. direct predator information

In Sect. 6.2.2, I identified a linear dependence of the evolutionary stable alignment strength on the flee strength and hypothesized that a balance between the benefit and cost of social information might be an explanation for this dependence. A prey can benefit from stronger alignment if it has no personal information about the predator's position. The benefit increases the quicker the alignment and therefore should increase with alignment strength. But if the prey is fleeing already, i.e. it has personal information on the predator position, then alignment to uninformed neighbors can hinder an escape. Therefore, I expect a balance between benefits and costs. In the following I show its semi-analytical approximation that reveals the observed linear dependence. Note that I do not want to claim that it is the only explanation for the linear dependence but a reasonable one.

The costs to align with uninformed prey, if the predator position is known, can be viewed as a deviation from the flee direction, i.e. the prey relaxes to an effective flee direction, which is the compromise between the mean direction of its neighbors and the flee direction Fig. A.5.

I will use the following assumptions:

- i) highly ordered: all neighbors are perfectly aligned with each other.
- ii) strong forces: the acting forces are strong such that the agents equilibrate quickly in the direction of the force.
- iii) constant forces: the flee-angle and the heading of the neighbors are not changing.
- iv) no noise: this will enable us to solve the problem analytically.

Consequently the change of the direction-angle of Eq. 3.1b can be reformulated to

$$\frac{d\varphi_i}{dt} = \frac{1}{v}\left(F_{i,\varphi} + \sqrt{2D}\xi\right) \tag{A.13a}$$

$$\approx \frac{1}{v}\left(F_{i,\varphi}\right) \tag{A.13b}$$

$$\approx \frac{1}{v}\left(\mu_{flee}\hat{f}_{flee} + \mu_{alg}[\langle\vec{v}\rangle_{N_i} - \hat{e}_{r,i}]\right)\cdot\hat{e}_{\varphi,i}. \tag{A.13c}$$

With $\langle\vec{v}\rangle_{N_i}$ being the mean velocity of all neighbors of agent i and $\hat{e}_{r,i}$ and $\hat{e}_{\varphi,i}$ are its parallel and perpendicular direction relative to its velocity, respectively. Without loss of generality I can permanently rotate the system such that $\varphi = 0, \forall t$. It simplifies the vector products since $\hat{e}_{r,i} = [1,0] = \hat{e}_x$ and $\hat{e}_{\varphi,i} = [0,1] = \hat{e}_y$. The angle α between $\vec{v}_i$ and $\langle\vec{v}\rangle_{N_i}$ behaves exactly opposite as φ (see Fig. A.5A) and its dynamics are reformulated to:

$$\frac{d\alpha}{dt} = -\frac{d\varphi}{dt} \tag{A.14a}$$

$$\approx -\frac{1}{v}\left(\mu_{flee}\hat{f}_{flee} + \mu_{alg}[\langle\vec{v}\rangle_{N_i} - \hat{e}_x]\right)\cdot\hat{e}_y \tag{A.14b}$$

$$\approx -\frac{1}{v}\left(\mu_{flee}f_{flee,y} + \mu_{alg}\langle\vec{v}\rangle_{N_i,y}\right)\ . \tag{A.14c}$$

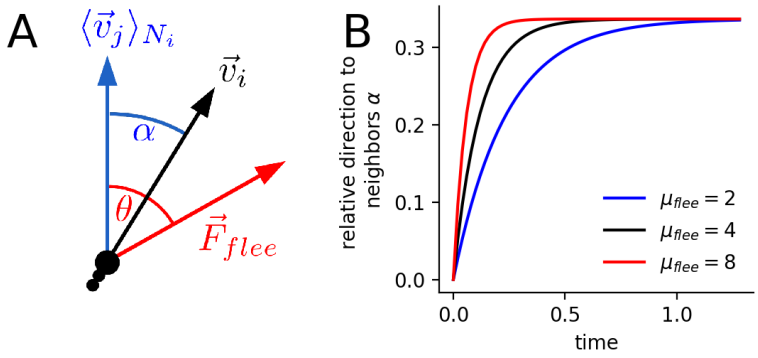

Figure A.5: Balancing social and private information via a directional compromise. A: Illustration of angle-vector-relations for variables used in Eq. A.14 and the following. The angle α is the angle between the mean velocity of neighbors $\langle \vec{v}_j \rangle_{\mathbb{N}_i}$ (blue arrow) and the velocity $\vec{v}_i$ of agent i (black arrow). The angle θ is the angle between the mean neighbor-velocity and the flee force $\vec{F}_{flee}$ (red arrow). **B**: Numerical-results of the relative direction to neighbors α using Eq. A.14. The initial conditions is $\alpha = 0$, i.e. the focal agent is perfectly aligned with its neighbors. The angle between mean neighbor velocity and flee force is $\theta = \pi/2$.

With $f_{flee,y} = \sin(\theta - \alpha)$ and by assuming perfect order and unit speed the mean velocity of neighbors is $\langle \vec{v} \rangle_{N_i} = 1 \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}$. Therefore, the change of α simplifies to:

$$\frac{d\alpha}{dt} \approx -\frac{1}{v} \left(\mu_{flee} \sin(\alpha - \theta) + \mu_{alg} \sin \alpha \right) \tag{A.15a}$$

$$\approx \mu_{flee} \sin(\theta - \alpha) - \mu_{alg} \sin \alpha. \tag{A.15b}$$

The fixed points are, as a sanity check, computed for the extreme cases $\mu_{alg} \gg \mu_{flee}$ and $\mu_{flee} \gg \mu_{alg}$ to $\alpha^\star = 0$ and $\alpha^\star = \theta$, respectively. There are in general four fixed points from which only one fulfills the criteria $\alpha^\star/\theta \in [0, 1] \forall \, (\mu_{flee} > 0, \ \mu_{alg} > 0, \ 0 < \theta < \pi/2)$:

$$\alpha^\star(\theta_s, \ \mu_{alg}, \ \mu_{flee}) = \arccos \frac{\mu_{alg} + \mu_{flee} \cos \theta}{\sqrt{\mu_{alg}^2 + \mu_{flee}^2 + 2\mu_{alg}\mu_{flee} \cos \theta}}. \tag{A.16}$$

Thus, $\alpha^\star$ is the effective flee angle with respect to the mean direction of the neighbors. The closer it is to the flee angle θ the smaller the cost of being aligned given the knowledge of the predator's position.

Now, I assume that individuals evolve such that they maintain $\alpha^\star(\theta_s)$ with respect to a specific θ_s. Thus, if the equilibration point $\mu^\star_{alg,evo}(\mu_{flee,evo})$ is known for a specific flee strength, which was used during the evolution $\mu_{flee,evo}$, the effective flee angle can be compute $\alpha^\star(\theta_s, \ \mu^\star_{alg,evo}, \ \mu_{flee,evo}) = \alpha^\star(\theta_s)$. If I assume that agents evolve such that the balance between alignment benefit and cost, manifested in the effective flee angle, is kept constant, then I can predict the evolutionary stable state $\mu^\star_{alg}$ for a given flee strength by reformulating Eq. A.16 to

$$\mu^\star_{alg} = \frac{\sin(\theta_s - \alpha^\star(\theta_s))}{\sin \alpha^\star(\theta_s)} \mu_{flee}. \tag{A.17}$$

The term $\frac{\sin(\theta - \alpha^\star)}{\sin \alpha^\star}$ does not depend on θ_s, which I confirmed numerically. Thus, the exact choice of θ_s is irrelevant and $\frac{\sin(\theta - \alpha^\star)}{\sin \alpha^\star}$ is only the slope that connects the origin and the one evolutionary stable state $(\mu^\star_{alg,evo}, \mu_{flee,evo})$ used to compute $\alpha^\star(\theta_s)$ as shown by the blue line in Fig. 6.6B.

Note that the equilibrium alignment strength $\mu^\star_{alg}$ above but close to the order transition is systematically lower than its predicted value, as seen for $\mu_{flee} \in \{2, 3, 4\}$ in Fig. 6.6B. This can be

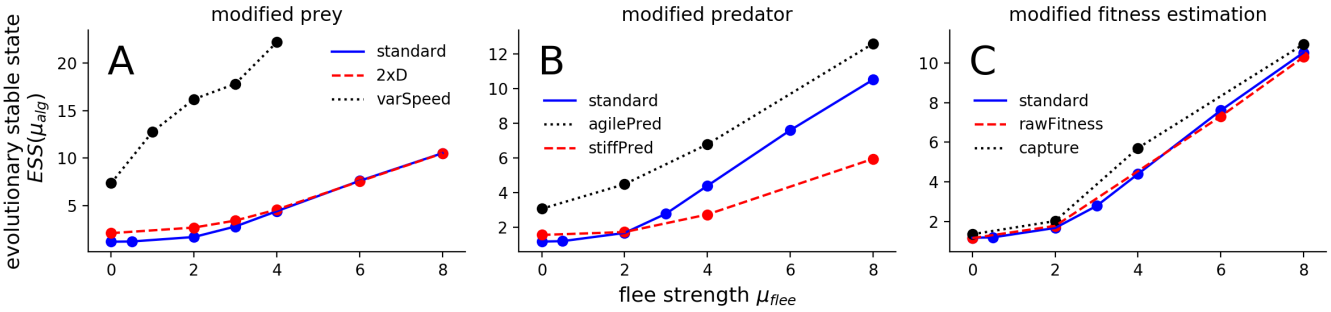

Figure A.6: Robustness analysis of evolution results. Evolutionary stable states of the alignment strength are estimated from the fitness gradient for different flee strength under slight variations of simulations parameters or predator attack implementation. The standard scenario (blue line) is compared to (**A**:) a prey population with varying speed that can avoid the predator additionally by acceleration (black dotted line), a prey population with an angular diffusion coefficient that is doubled compared to the standard case (red dashed line), (**B**:) a stiff predator that turns less quick (black dotted line) and an agile predator that turns quicker (red dashed line) than the predator in the standard case. (**C**:) a non-binarized fitness estimate (red dashed line) in which the preys fitness is not defined by captures but by the accumulated probability to get caught, a fitness estimate based on captures during the simulation (black dotted line),

explained by a small signal due to the low flee strength, because the system relaxes faster the greater the flee strength μ_{flee} (see Fig. A.5B). An alternative explanation is that the spatial selection due to strong self-sorting dominates at the transition. This explanation is also in agreement with the ESS for low flee strength ($\mu_{flee} = 0.5$) being identical to the one with no flee strength at all ($\mu_{flee} = 0$).

A.6 Robustness analysis of evolution results

To ensure that the results are robust the evolution is repeated (Fig. A.6) with (i) modified prey properties, i.e. changing the angular diffusion coefficient and introducing variable speed, (ii) a changed predator behavior, i.e. its agility, and (iii) changes in the evolutionary selection mechanism, e.g. by a predator that captures the prey during the simulation (and not that prey are declared as dead after the simulation).

A.6.1 Prey modifications

The change in angular diffusion from $D = 0.5$ to $D = 1$ shifts the order-transition to a larger mean alignment strength of $\mu_{alg,c} \approx 1.6$ and therefore also increases the lower bound for the ESS. This is visible in larger ESS for small flee strength (compare dotted red with blue line in Fig. A.6A). For larger flee strength the results are nearly identical suggesting that the mechanism defining the ESS remains unchanged with respect to the standard scenario (corresponding to parameters listed in Tab. 3.1).

If the speed of the prey is not constant but can change according to social forces, the equations of motion (Eq. 3.1) change to

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i \quad \text{with } \vec{v}_i = v_i[\cos\varphi_i, \sin\varphi_i] \tag{A.18}$$

$$\frac{dv_i}{dt} = \beta(v_0 - v_i) + F_{i,v}(t) \tag{A.19}$$

$$\frac{d\varphi_i(t)}{dt} = \frac{1}{v}\left(F_{i,\varphi}(t) + \sqrt{2D}\xi(t)\right) \tag{A.20}$$

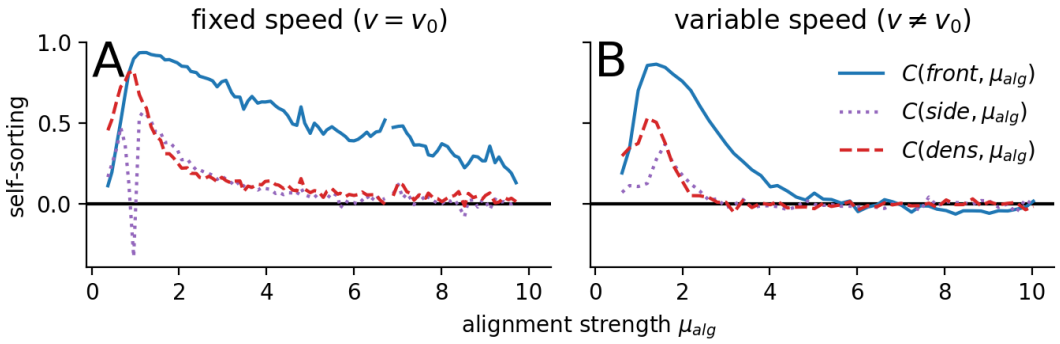

Figure A.7: Self-sorting with and without fixed speed. Self-sorting quantified via the Pearson correlation between the individual alignment parameter μ_{alg} and the average relative position of the individuals (relative front-, side- or density-location as described in Sect. 6.1.3). **A**: If prey agents respond only by changing their direction but not their speed (fixed speed), self-sorting persists also in highly ordered regions. **B**: If prey agents can change their speed (variable speed), self-sorting vanishes for $\mu_{flee} \leq 6$.

with $F_{i,v}(t) = \vec{F}_i \cdot \hat{e}_{h,i}$ as the projection of the social force of prey i on its heading direction $\hat{e}_{h,i}$ and β as the relaxation coefficient, which is set in the following to $\beta = 4$. A value of $\beta = 4$ prevents the school to relax into a non-moving phase that exists for lower values of β (Großmann et al., 2012). In this non-moving state the speed of the prey fluctuates around zero. Additionally an upper bound for the preys speed is set, corresponding to eighty percent of the predator's speed $v_{max} = 0.8v_p$. Non-fleeing prey ($\mu_{flee} = 0$) evolve to significant larger values compared to the standard scenario (compare dashed black with blue line in Fig. A.6A). The ESS for non-fleeing prey ($\mu_{flee} = 0$) coincides with the zero-crossing of the front-sorting (Fig. A.7). Not only is the ESS of the non-fleeing prey at larger values due to a different self-sorting but also is the ESS much more sensitive to changes in the flee strength (compare slope of dashed black with blue line Fig. A.6A). This steeper increase is explainable with an additional social cue, the increased speed of fleeing-prey, which was not present in the standard-scenario and goes in hand with existing studies (Lemasson et al., 2009, 2013).

A.6.2 Predator modifications

I repeated the simulations with (i) a stiffer predator that turns slower and (ii) a more agile predator that turns faster compared to the predator in the standard scenario. The different turning ability are implemented by modifying the pursuit strength μ_p to $\mu_p = 1$ for the stiff and to $\mu_p = 3$ for the agile predator.

The effect of using the stiff predator is negligible for low flee-strength, probably because the order-disorder transition acts as lower bound for the ESS due to the explained maximum in assortative mixing and resulting subpopulation selection. However, for larger flee strength, e.g. $\mu_{flee} \in \{4, 8\}$ in Fig. A.6B, the ESSs are lowered compared to the standard scenario. This can be explained by the missing feedback between the reaction of the prey and the trajectory of the predator: in the standard scenario the predator heads for the closest prey (actually the direction weighted by distance). Thus, if certain prey are good at avoiding the predator, they have an additional fitness benefit because the predator pursues less well-avoiding prey.

Consequently, the trajectory of the more agile predator should have a stronger feedback with the reaction and lead to larger evolutionary stable alignment strength. This is in fact observed (compare dotted black with blue line in Fig. A.6B). Despite the explainable variations due to the predator modifications, the general findings, i.e. that the ESSs are in the ordered phase and increase with increasing flee-strength, are robust.

Attack direction and simulation time

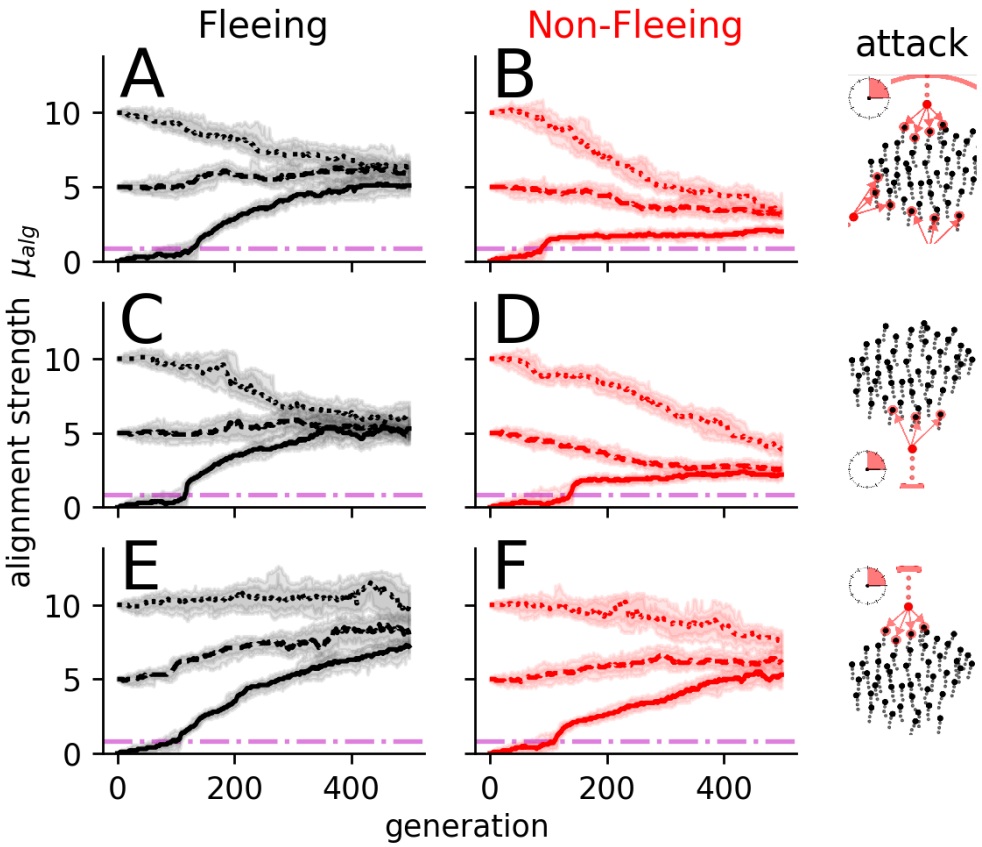

Figure A.8: Evolution under different initiation angle and simulation time. The predator attack duration of $T_{simu} = 40$ is 3 time shorter than to the standard parameters (Tab. 3.1). The predator is initiated at random position around the shoal (**A, B**), behind (**C, D**) and in front of the shoal (**E, F**) as illustrated at the right side of each row. The prey shoal is initiated at three different alignment strength at generation zero. The evolution is simulated for fleeing ($\mu_{flee} = 4$, **A, C, E**) and non-fleeing ($\mu_{flee} = 0$, **B, D, F**) prey. All panels: the magenta dash dotted horizontal line marks the order disorder transition.

The standard predator is initiated at a random angle relative to the shoal heading direction and the simulation lasts for $T_{simu} = 120$ time units. If the simulation time is decreased to a third of the standard time ($T_{simu} = 40$) the ESS remains about the same for fleeing and non-fleeing prey (Fig. A.8A, B). This is also true if the predator appears always behind the shoal (Fig. A.8C, D). However, if the predator appears in the front, the ESS shifts to larger alignment strength for fleeing and non-fleeing prey (Fig. A.8C-F).

The fact that the ESS of non-fleeing prey is affected suggests that it is a structural cause. This effect is very robust and even stronger for an alternative predator implementation: the predator follows the center of mass (COM) and continues straight if it gets close to the COM (in detail discussed in Sect. A.7). In this case the trajectory of the predator should not differ much between front and back initialization. The reason for the different ESS is the "border dilution effect", a consequence of the Voronoi implementation, and in detail explained in Sect. A.7.1. In short, if the predator appears outside the shoal, it has more frontal Voronoi neighbors as in the bulk and therefore also a larger number of possible targets. The predator selects equally among all possible targets (irrespective their distance). This reduces the catch probability according to Eq. 3.13, it gets diluted. Thus, the individuals where the predator is initiated have a diluted risk and therefore

a fitness benefit. Combined with the front-sorting (Fig. 6.4C) it explains the shift of the ESS for evolution under the frontal initiated predator attacks.

A.6.3 Selection modification: capture during simulation

In the standard simulations, prey are not captured but a fixed fraction of them with the largest accumulated probability to get caught is declared as captured after the simulation. This means that no prey is removed during the simulation. It reduces stochasticity of the fitness estimate but can be considered as unrealistic. If prey are removed during the simulation based on their current probability to get caught and on the predator's attack rate, the evolution results remain unchanged (compare dotted black with blue line in Fig. A.6C). Hereby is the attack rate γ_a adjusted at each generation g such that the mean capture rate $\langle \gamma_c \rangle$ matches the initially set attack rate $\gamma_a(g = 0)$:

$$\gamma_a(g+1) = \gamma_a(g) * \frac{\gamma_a(0)}{\langle \gamma_c(g) \rangle}. \tag{A.21}$$

This ensures a constant evolutionary pressure.

The attack rate parameter can be abandoned if the fitness is not estimated by the captures but by the negative accumulated probability to get caught. This modification does not alter the ESS identified in the main text at all (compare dashed red with blue line in Fig. A.6C).

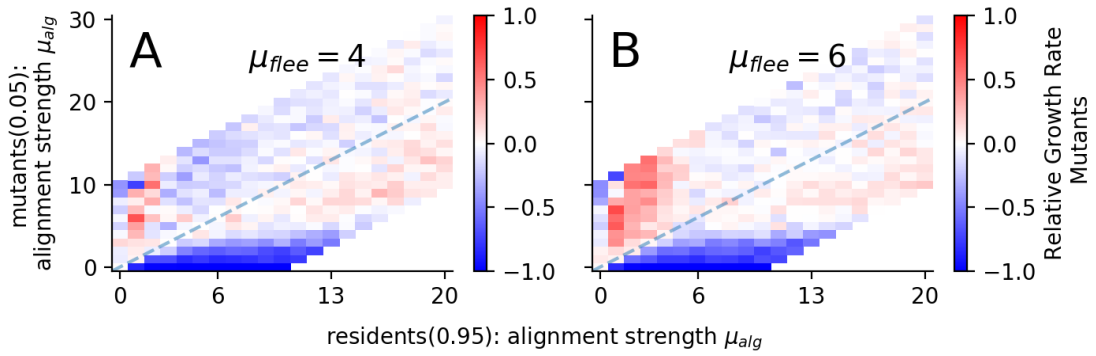

Figure A.9: Invasion fitness landscape. Created for a population consisting of five percent mutants and 95 percent residents for prey with a flee strength of $\mu_{flee} = 4$ (**A**) and $\mu_{flee} = 6$ (**B**). Each box represents the relative growth rate of mutants averaged over $N_{sam} = 21$ sample simulations. The diagonal dashed line marks where the population is fully homogeneous, i.e. where the mutants have the same parameters as the residents.

Another alternative is to fully abandon the evolutionary algorithm and deduce the ESSs from the invasion fitness landscape (Waxman and Gavrilets, 2005). In this approach a homogeneous population (residents) is invaded by a small number of invaders (mutants). The relative growth rate γ_g is computed as

$$\gamma_g = \frac{N_{m,\text{offspring}}}{N_m}, \tag{A.22}$$

i.e. the number of mutants in the offspring generation $N_{m,\text{offspring}}$ relative the number of initial invaders N_m. The ESS is located where no mutants can invade the residents. The invasion fitness landscape for two different flee strength $\mu_{flee} = [4,\ 6]$ is shown in Fig. A.9. The ESS approximated by this technique corresponds roughly to the ones reported in Fig. 6.6. However, it is not as straightforward to identify the ESS. It is likely more straightforward with finer scans and larger sample rates.

A.7 Evolution under a straight predator attack

The non-fleeing prey are introduced as a null-model in which individuals evolve only based on their relative position. Thus, non-fleeing prey represents the spatial component and are therefore easier to interpret. However, even the simplest simulations using non-fleeing prey are difficult to understand. Consider for example a predator that attacks from the rear vs. from the front. In both cases the predator heads for the center of mass, has therefore roughly the same path through the collective with the difference that one predator starts where the other ends. This difference alters the evolution path: for the front attacking predator positive above $\langle \mu_{alg} \rangle = 6$ while for the rear attacking predator its negative (Fig. A.10). One would expect a similar evolution path because the difference between the simulations seems marginal.

This discrepancy is explainable by border-effects appearing due to the chosen form of the predator's selection probability p_{select}.

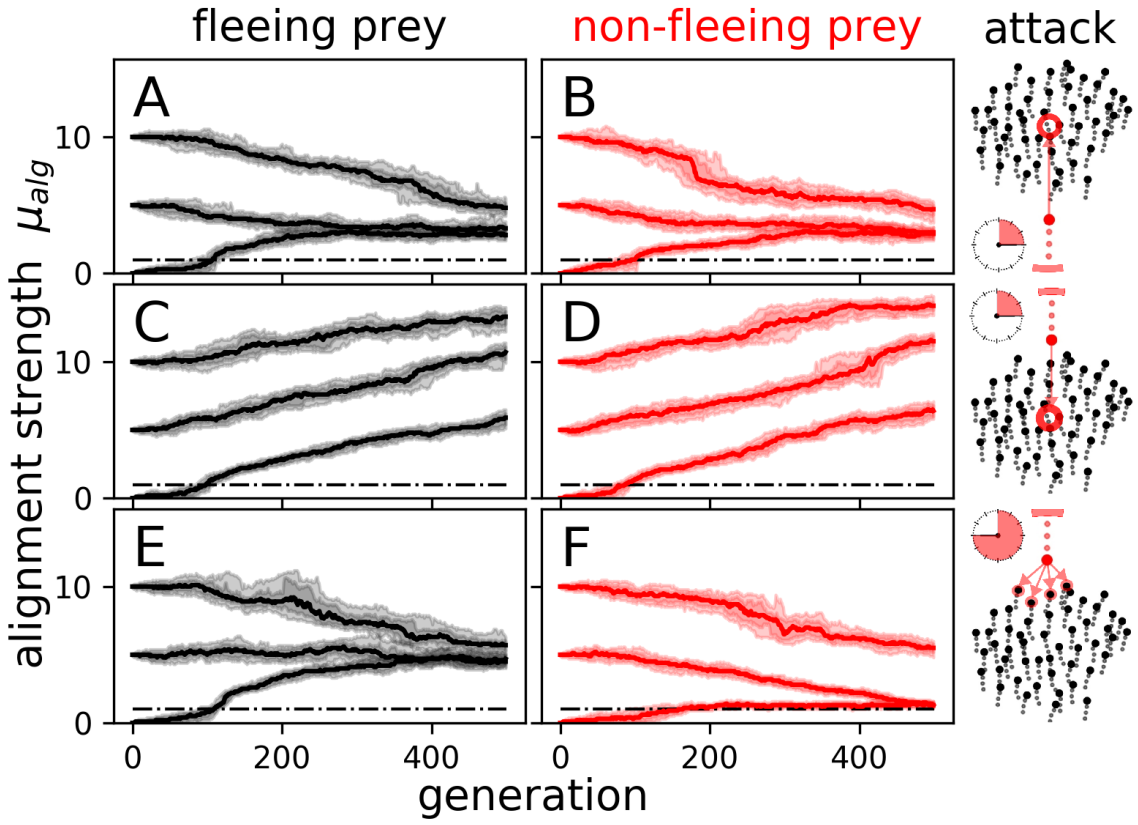

Figure A.10: Evolution under a straight attacking predator initialized in front or rear. Evolution results if a predator attacks from the rear (**A, B**) and front (**C, D**) and follows the center of mass (COM). After the predator reaches the COM it heads straight and the simulation is stopped when no prey is in its front. For comparison, the standard predator, which follows its the weighted direction to its closest targets, is simulated and attacks from the front (**E, F**). For every attack mode fleeing ($\mu_{flee} = 4$, **A, C, E**) and non-fleeing ($\mu_{flee} = 0$, **B, D, F**) prey evolve. Each attack scenario is initiated at $\langle \mu_{alg} \rangle (g = 0) = [0, 5, 10]$.

A.7.1 Border dilution effect

The reason for the difference in evolution of front- vs. rear-attacked non-fleeing prey is a "border dilution effect" caused by the border-structure, the Voronoi implementation and the predator's

selection probability p_{select}. The predator enters the swarm for rear-attacked prey from the back and for front-attacked prey from the front. If the catch probability p_{catch} is decreased at the entry of the predator in the shoal, by an effect explained below, the evolution of front- (rear-) attacked prey would favor frontal (rear) individuals. Because agents at the front tend to have a higher alignment tendency the decreased catch at the shoal-entry would result in the observed discrepancy between front- vs. rear-attacked prey.

This border dilution/confusion is explained by the almost linear alignment of the prey at the border as seen by a typical snapshot in Fig. A.11A. In contrast to a predator in the bulk, shown in Fig. A.11B, the number of targeted prey is larger and the distance to predator varies more and is on average larger as in the bulk. Since the predator selects with equal probability among its targets (Eq.3.11) the increased number of prey targets and their increased mean distance to the predator effectively lowers the mean catch probability of the predator at the border region.

If the former description is correct, the discrepancy between front- vs. rear-attacked could be increased by letting the predator start closer to the swarm. This would expose the agents at the border shorter to the predator and should therefore decrease their probability to be captured even further but leaving the behavior in the bulk untouched. This expected increase in difference between the evolution runs is verified by simulation (Fig A.11C). Here, the fitness gradient $\Delta fit(\mu_{alg})$ of rear- and front-attacked prey is estimated for the normal attack distance of $r_p = 4.5$ and the reduced attack distance of $r_p = 1$.

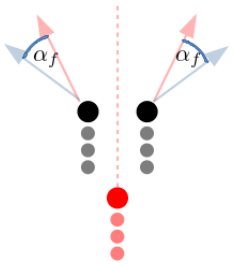

Figure A.12: Alternative flee mechanism (blue arrow) alters the direct repulsion (red arrows) by a fixed extra flee angle α_f.

In contrast, given that the border dilution/confusion effect causes the difference, the difference should decrease if the predator's selection probability p_{select} is modified such that it is one for the closest prey and zero otherwise, i.e. if the predator follows the closest target. If the simulation is repeated with this modified selection probability the difference in fitness gradient $\Delta fit(\mu_{alg})$ between rear- and front-attacked prey almost vanishes (Fig. A.11D).

Since the estimated fitness gradient assumes a fixed standard deviation of the μ_{alg}-parameter distribution in the population, it is, especially when close to zero, not an equivalent representation of the evolution. Therefore the original evolution of the non-fleeing prey attacked by a straight-predator is repeated with the modification of a predator selecting the closest prey as target (Fig. A.13A). Interestingly, the effect is inversed, i.e. the front-attacked prey evolve to a lower ($\langle\mu_{alg}\rangle \approx 4$) while the rear-attacked evolve to a larger mean alignment strength ($\langle\mu_{alg}\rangle > 6$). Note that this inversion is not in conflict with the approximated fitness gradient from Fig. A.11D because the gradient is close to zero and the parameter-distribution would therefore change. The inversion can be explained by the initial predator distance to the collective as already mentioned related to Fig. A.11C. If we repeat the evolution with a shorter initial predator distance $r_p = 4.5 - > 1$ (Fig. A.13B) the results are comparable to the original simulation-case. Therefore, if we choose an appropriate predator distance r_p both simulation would eventually evolve to the same mean alignment strength. In summary, the border-dilution/confusion effect is the reason for the front- and rear-attacked non-fleeing prey to evolve to different mean alignment strengths.

A.7.2 Evolution to bimodal distributions: differentiation?

In this section the straight attacking predator was always initiated from the front or the rear but never randomly or from the side. This is done because both other cases, i.e. random and side attacks, can lead to a bimodal distribution caused by the front-sorting of individuals. Interestingly, one peak of the distribution is at extremely low alignment strength $\mu_{alg} \approx 0$ and the other is constantly evolving to larger alignment strengths and corresponds to a strongly ordered phase. However, it is an artefact of the implementation because the disordered subgroup of the collective is at the very back of the swarm and can not keep up with the much faster ordered subgroup in the front. The shoal would split if the equilibration time would be longer, but like that the shoal

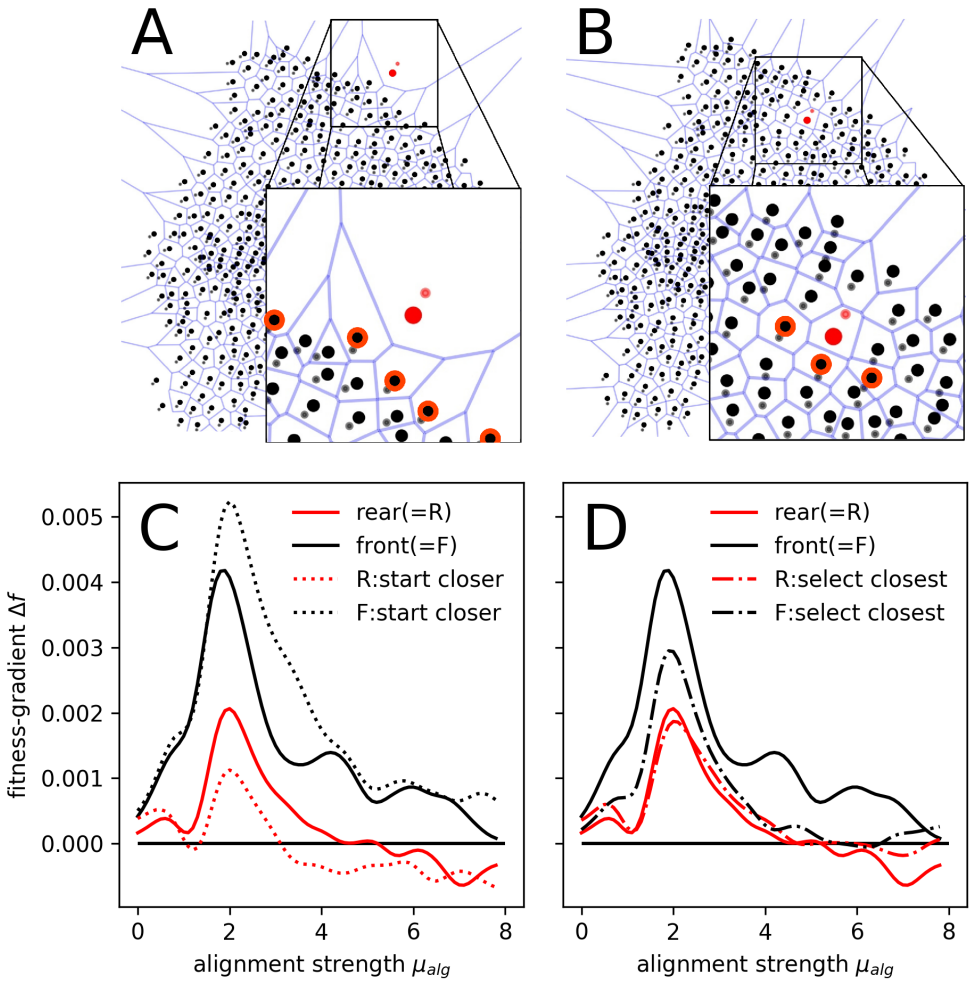

Figure A.11: Border dilution effect: illustration and fitness gradient manipulation. Simulation snapshots of a predator, in red, just outside (**A**) and inside (**B**) the swarm. The insets show magnifications of regions close to the predator with the frontal Voronoi neighbors of the predator marked by red circles. **C,D**: estimated fitness gradient Δf for different simulation scenarios. The gradient is estimated by a sample of 100 simulations per mean alignment strength μ_{alg} with a standard deviation of $\sigma_{alg} = 0.5$ for its parameter distribution. The original front- and rear-attacked simulations are compared with modified versions in which the only difference is a closer starting point of the predator to the swarm (**C**) or an alternative selection mechanism to Eq. 3.11 in which only the closest prey is selected as possible target (**D**). Red line represents rear- and black front-attacked swarms, the original simulations are solid lines, the ones starting closer to the swarm are dashed and the simulations with the alternative selection mechanism are dash-dotted.

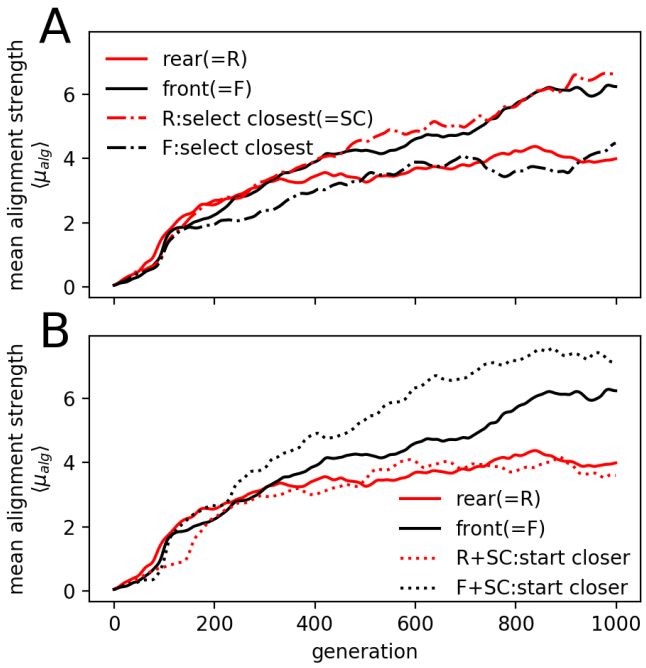

Figure A.13: Evolution on non-fleeing prey under modified straight predator attacks. In contrast to the standard predator (follows its distance weighted targets) the straight attacking predator pursues the center of mass (COM) of the shoal and continues straight if it is sufficient close to the COM. The evolution with the other parameters identical to the standard case (Tab. 3.1) is shown with a solid line (**A, B**). Evolutions with a predator selecting the closest frontal prey as target (dash dotted line, **A**) and additionally where the predator appears closer to the collective (dotted line, **B**). Red and black colored lines are rear- and front-attacked prey, respectively.

is very elongated. Consequently, a straight attacking predator initiated on the side of the shoal would only target agents with intermediate alignment strength and therefore increase grant fitness benefits to the agents in the very front and the very back of the shoal.

This is mentionable, because it is a possible pitfall for anyone who would like to continue this study. It seems appealing to use such a simple predator attacking mechanism as the "straight" type, but be aware of possible problems as the artificial bimodality reported here.

A.8 Optimal flee-direction

A direct repulsion might not be the best mechanism to avoid a predator. Here, I show another alternative mechanism that is based on the normal repulsion but adds an extra angle α_f away from the future predator trajectory (assuming no change in the predator's velocity). This mechanism is illustrated in Fig. A.12. If the distance to the predator is computed for different flee angles α_f, it turns out that for a specific predator speed and prey parameters (different from standard parameters, because it is based on an older model implementation) the prey build up the largest distance to the predator with $\alpha_f = \pi/4$ (Fig. A.14).

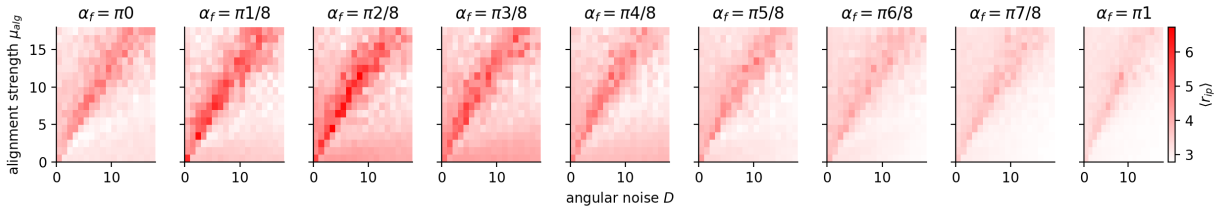

Figure A.14: Alternative flee angle exploration. Mean distance to the predator r_{ip} for different flee angles α_f. Note that the units of alignment strength and angular diffusion coefficient are arbitrary.

B Attention trade-off in zebrafish selection lines

B.1 Data processing

B.1.1 Validating high acceleration trajectories

In Sect. 9.2 the id-Tracker output is filtered by excluding extreme jumps according to acceleration criteria. The resulting acceleration distribution still contains a considerable amount of large positive values. To ensure that the criteria is appropriate the trajectory of the 13 largest accelerations are visualized and checked for inconsistency the a normal trajectory (Fig. B.1). Almost all high-acceleration trajectories (Fig. B.1B-M) show a startling event and are therefore valid. A startling event is marked by an extreme acceleration, however the individuals speed decreases after this acceleration continuously. Only one trajectory shows a clear tracking error (Fig. B.1A) visible by the sudden change to the previous speed after a moderate acceleration period of 1 frame.

B.1.2 Smoothed trajectories

To estimate the smoothening width of the Gaussian kernel a randomly selected trajectory of 2000 frames was analyzed. The burst-events were counted manually, which is easy because a burst is normally accompanied by a turning event. The numbers of bursts were compared to the automatically detected number of burst-events, which are the number of periods of positive acceleration. Results are displayed in Fig.B.2

B.2 Parameter estimation

B.2.1 Estimation of social zone ranges

In Sect. 11.1.4 it is described how the different ranges of the social zones are estimated. I mentioned there only shortly that the parameter-scans around the identified optima revealed that the optima corresponds to a two-zone model, i.e. where no orientation zone exists but only repulsion and attraction. Here I present the results in detail in Fig. B.4. The most important panels are C, F, I in which it is apparent that the angle difference between predicted and experimentally observed burst-direction is smallest if the orientation range r_o equals the repulsion range r_r, i.e. fish never align with each other consistently across selection lines.

B.2.2 Details on optimization runs

Here, more detailed results on the parameter setting of burst force F and probability to follow social cues P_{soc} with the CMA-ES are shown. In Fig. B.5A-F the final parameters, set by the CMA-ES, are displayed. Each of the 4 results represents a different initialization (indicated by the lines). The fact that they differ only very little shows that the optimizer found a global minimum (and not a local one).

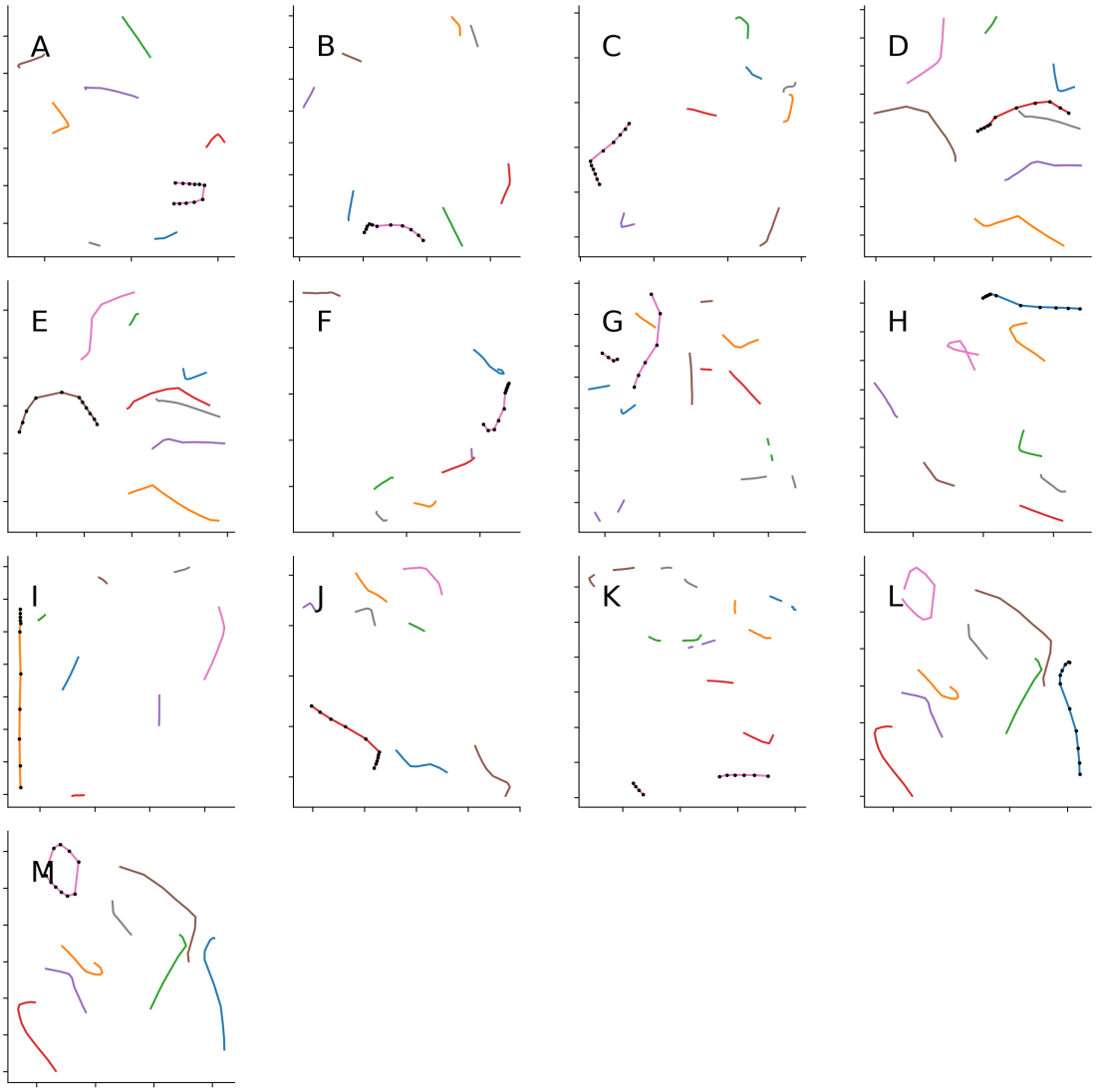

Figure B.1: Trajectory of the individual with an extreme acceleration, with its position marked by a dot at each frame. The other 7 individuals are also displayed to ensure that no identity switches happened. Each color represents the trajectory of one individual. Five frames before and five frames after the extreme acceleration are displayed.

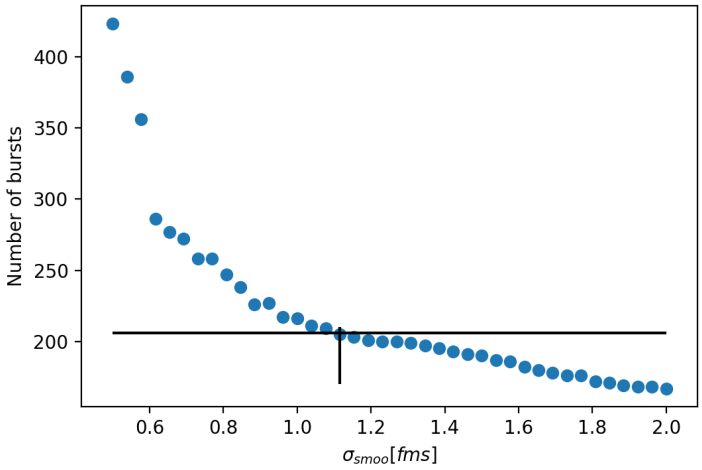

Figure B.2: Number of burst events of a randomly selected trajectory spanning 2000 frames. The horizontal black line marks the number of manually counted burst events. Blue dots mark automatic counts of burst events on a smoothed trajectory using a moving window with a Gaussian kernel of varying standard deviation σ_{smoo}. The vertical black line indicates the standard deviation $\sigma_{smoo} = 1.13fms$ where the number of automatically detected bursts equals the manually counted one in an interval of 2000 frames.

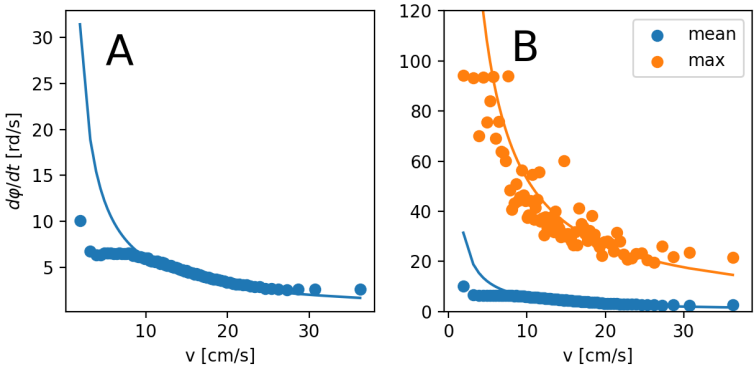

Figure B.3: Angular change $d\varphi/dt$ depending on the velocity v. The blue dots (**A, B**) represent the mean values while orange dots (**B**) the maximum observed value for the binned data. The lines are not fitted to the data but serve as a comparison to the inverse proportionality. The blue line is $f(v) = 6/v$ and the orange line is $f(v) = 530/v$. The resemblance of the data to the lines supports the model assumption in Eq. (10.2c).

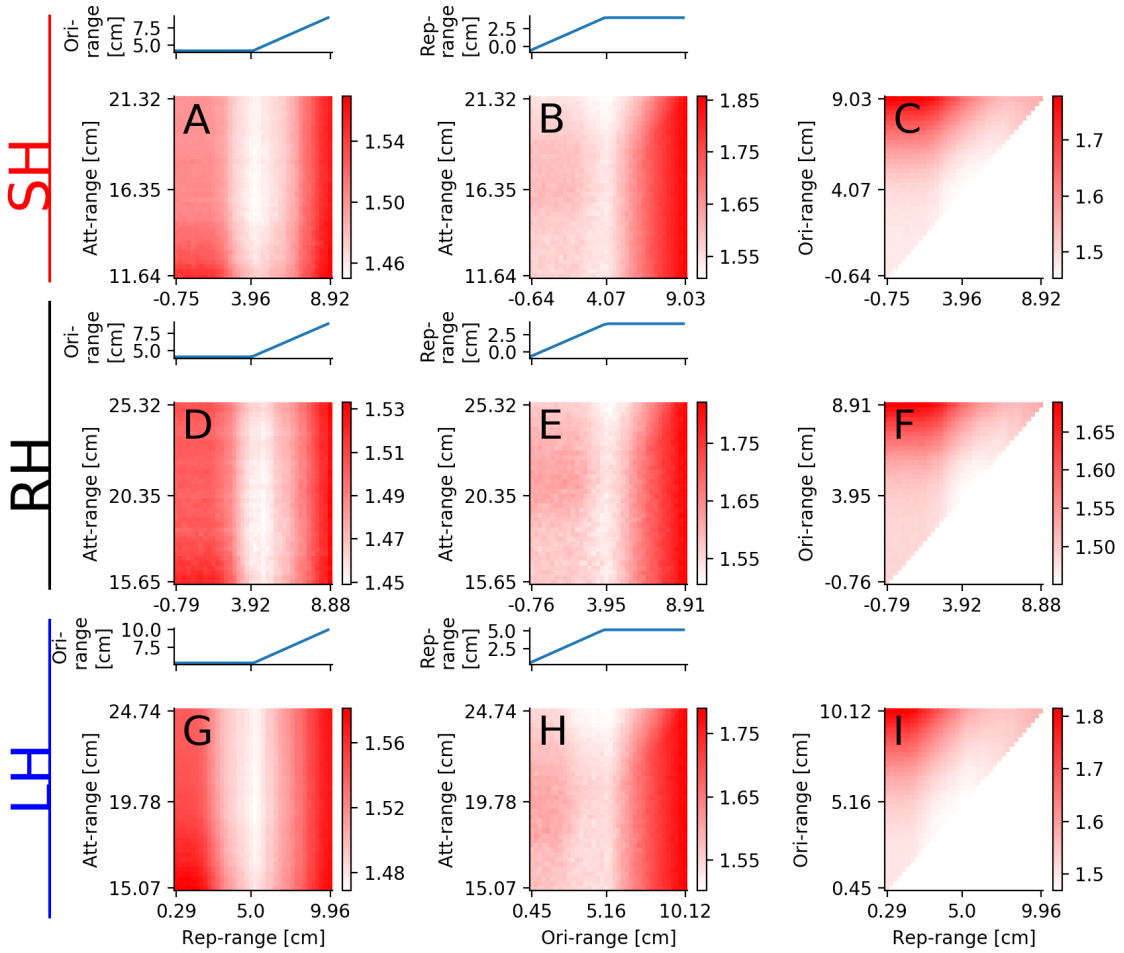

Figure B.4: Angle difference between predicted and actual direction after burst is color-coded. The difference is analyzed around the optimal values of the repulsion, orientation and attraction ranges. The line-plots above the color-plots show the remaining range which was kept constant if possible but needed to be in- or decreased to ensure the inequality $r_r < r_o < r_a$. In the first (**A, B, C**), second (**D, E, F**) and third (**G, H, I**) row is data used from burst-constellations of the small-, random- and large-harvested selection line, respectively. We varied the repulsion and attraction range in the first column (**A, D, G**), the orientation and attraction range in the second (**B, E, H**) and the repulsion and orientation range in the third column (**C, F, I**).

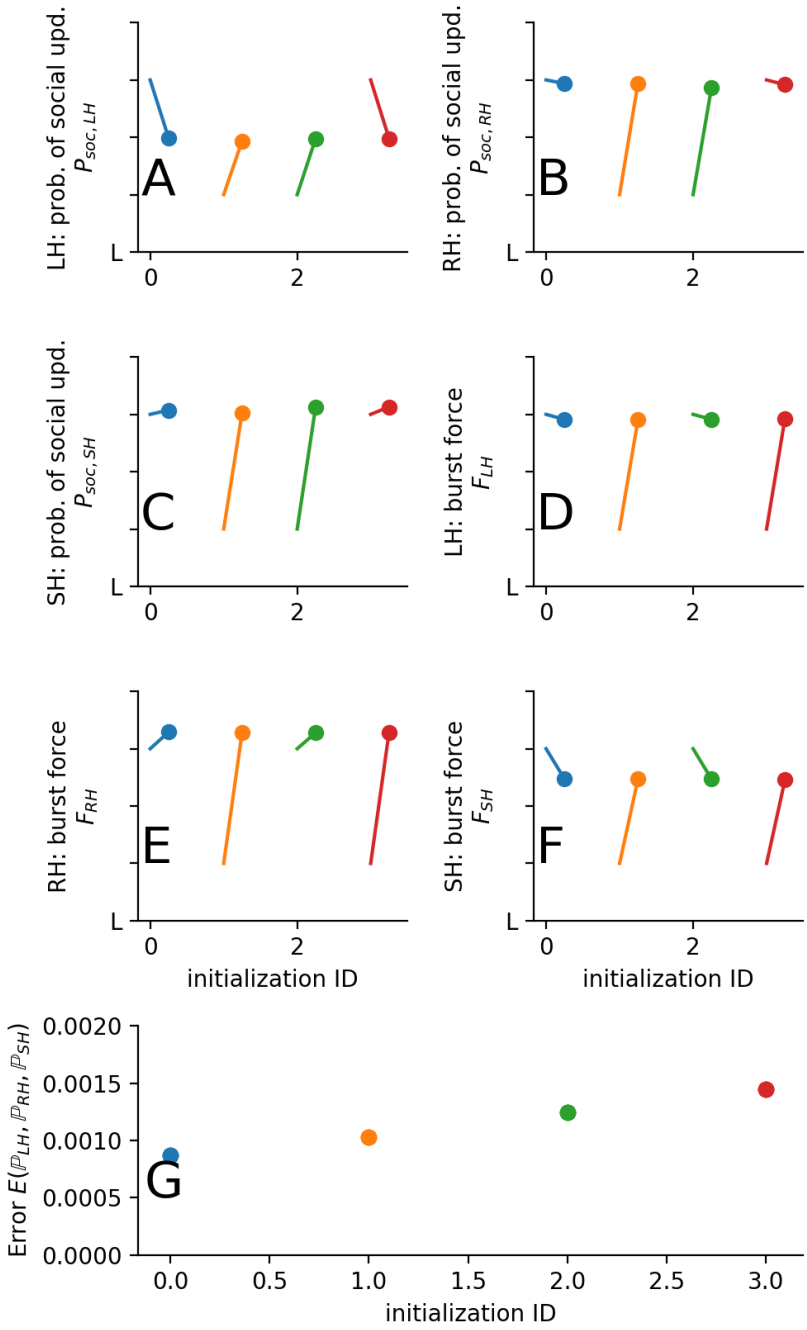

Figure B.5: Details on optimization outcomes for the selection line model representations. Initial and final parameters (**A-F**) together with final error (**G**) according to Eq. 11.5 for the 4 different initialized optimization runs. The initialization are sorted such that the one with the best outcome, i.e. with the smallest error, is first. Colors mark specific initialization ID and are consistent across sub figures. The circles (**A-E**) mark the final parameter while the other end of the line marks the initial setting.

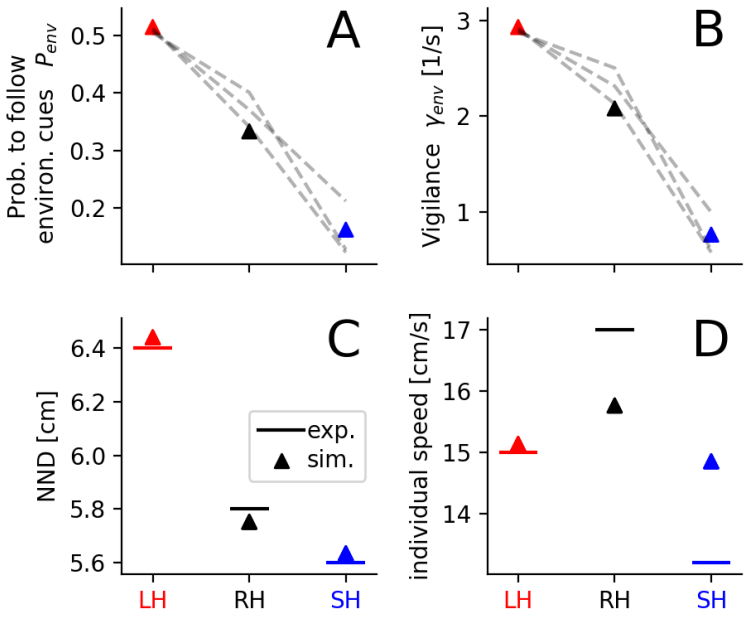

Figure B.6: Fitted model representation of the different selection lines with same burst force F. In order to resemble the experimental observations a probability to follow environmental cues (**A**) was set for each selection line and a **common burst force** F (not shown) was fitted. The vigilance γ_{env} is the product of P_{env} and the burst rate γ (**B**). The average nearest neighbor distance (NND, **C**) and the average individual speed (**D**) are emergent properties of the model and were used to quantify how well the parameters (P_{env}, γ) reproduce the experimental observations. Triangles represent the model parameters or the simulation outcomes of the parameter set with the best match to the experiment. Dashed lines (**A, B**) represent a parameter set of a different initialization, and therefore show the robustness of the best matching parameter set (triangles). The horizontal solid lines (**C, D**) represent the experimental values. The selection lines are indicated by different colors (red=LH: large-harvested; black=RH: random-harvested; blue=SH: small-harvested).

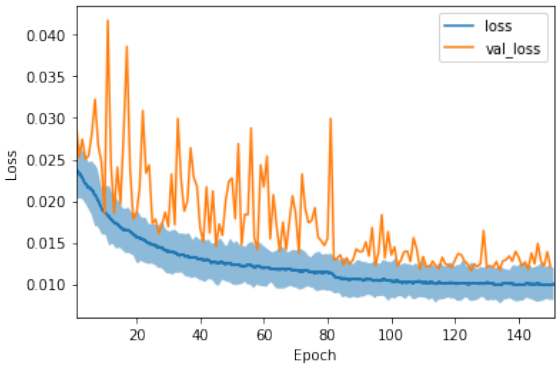

Figure C.1: Loss and validating loss of CNN *fishNet*

C Collective escape waves of P. sulphuraria

C.1 Box size selection for meso-scale wave speed estimation

For the meso-scale analysis the space is coarse grained via boxes of length $l_b = 100px$. This size selection represents a compromise between the number of the data points (spatial resolution) and the resolution of the wave speed. From the macro-scale analysis (Sect. 15.2) we know that the average wavespeed is $\langle v_w \rangle \approx 1.5m/s$. With a box length l_b and a frame difference $\Delta f = 1$ in the activation time of opposite neighbors of the focal box the detected wave velocity is

$$v_{box}(\Delta f) = \frac{2 \cdot l_b \cdot fps}{SF} \ . \tag{C.1}$$

Here, $fps = 50\text{frames}/s$ is the framerate of the camera and $SF = 1500px/m$ is the scale factor from pixel to meter (known from the homography/rectification of the video). Therefore, the fastest detectable wave is $v_{box}(1) = 6.67m/s$. This speed is about four times the average expected speed but the second fastest speed is $v_{box}(2) = 3.3m/s$ and the third is $v_{box}(3) = 2.2m/s$. That means the resolution gets better the farther the speed is away from the detection limit $v_{box}(1)$. The closest possible detectable speeds to $\langle v_w \rangle$ are $[2.2, 1.67, 1.3, 1.1]m/s$. This is far from fine but an agreeable compromise because otherwise we would lose spatial resolution.

C.2 Optimization artifacts for the wave speed and direction

In the main text the discrete nature of the approximated speed and direction of the collective escape where observed in Figs. 16.11, 16.9. The explanation for these discrete distributions is most likely the optimizer which tries to find the best planar wave which minimizes the error between predicted activation time and measured activation time (Eq. 16.4). I hypothesized that the optimizer selects a pair of opposite neighbors of the focal box which are in best agreement with the planar wave approximation. If the speed and direction of the waves are plotted together, this explanation is strengthened (Fig. C.2). That means, you see clearly that the velocity peaks are at multiples of $\phi_w = \pi/4$. Additionally the peaks at $\phi_w = \pm[\pi/4, 3/4\pi]]$ are larger than the peaks at $\phi_w = \pm[0, \pi/2, \pi]$. This happens because the latter directions ($\phi_w = \pm[0, \pi/2, \pi]$) corresponds to the side and front-back neighbor boxes which are closer to the focal box. The other directions ($\phi_w = \pm[\pi/4, 3/4\pi]]$) corresponds to the diagonal neighbors which are farther away and the wave needs to be faster to travell the distance in one frame. Thus, the pattern is a combination of the discrete neighbor box implementation and the discrete time (due to video recordings).

If the above explanation is true the I can even compute the theoretical expected pattern. Assume that from the focal box A the neighbor B and its opposite neighbor C corresponds best to the planar wave. In the following derivation I ignore C, but since C is opposite of B its relative location is $\vec{r}_{AC} = -\vec{r}_{AB}$ and the whole argumentation below holds. The activation time of box A and its neighbor B are $t_A = 0$, t_B. Thus, the speed of a planar wave with direction $\hat{v}_w = [cos(\phi_w), sin(\phi_w)]$ to reach B from A in time t_b is

$$v = \frac{(\vec{r}_B - \vec{r}_A) \cdot \hat{v}_w}{t_B} \tag{C.2}$$

$$= \frac{\vec{r}_B \cdot \hat{v}_w}{t_B} \ . \tag{C.3}$$

Without loss of generality we can rotate the system such that $\vec{r}_B$ is on the x-axis which leads to

$$v = \frac{r_B \cos \phi_w}{t_B}. \tag{C.4}$$

which is the dependency we observe in the data (Fig. C.2).

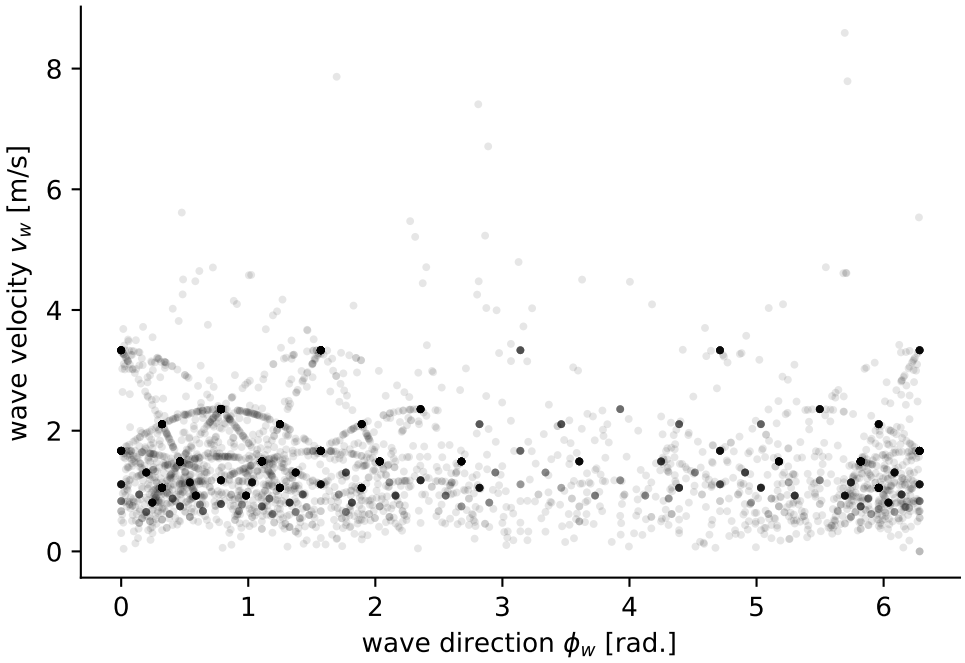

Figure C.2: Pattern in speed and direction of escape waves. The wave speed and its direction are shown as raw data points.

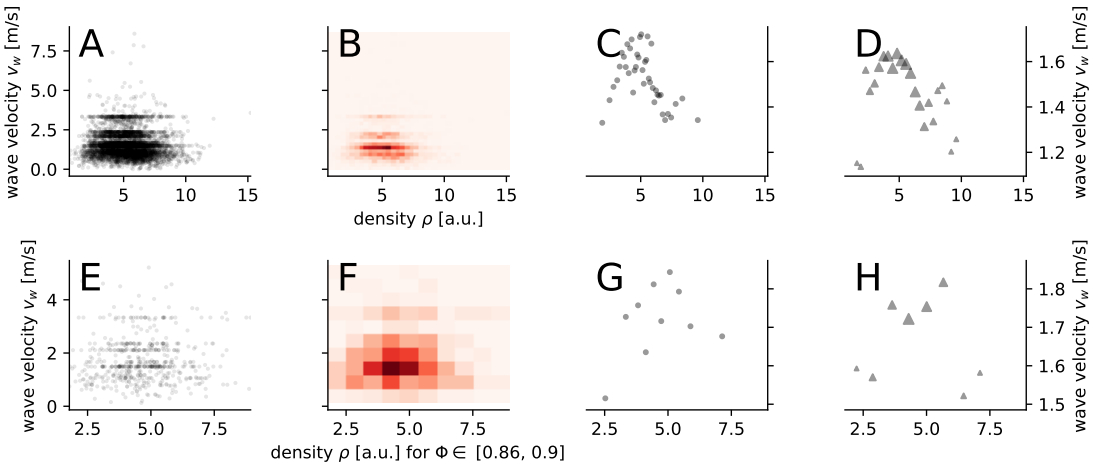

Figure C.3: Dependence of wave velocity v_w on the **density** of the full data set (**A-D**) and on densities corresponding to boxes with a polarizaton $\Phi \in [0.86, 0.9]$ (**E-H**). **A, E**: raw data points. **B, F**: 2D histogram of the raw data. **C, G**: accumulated data with each data points representing the same number of raw data points (equal weights). **D, H**: binned data in equidistant bins with the size of the triangles representing its weight.

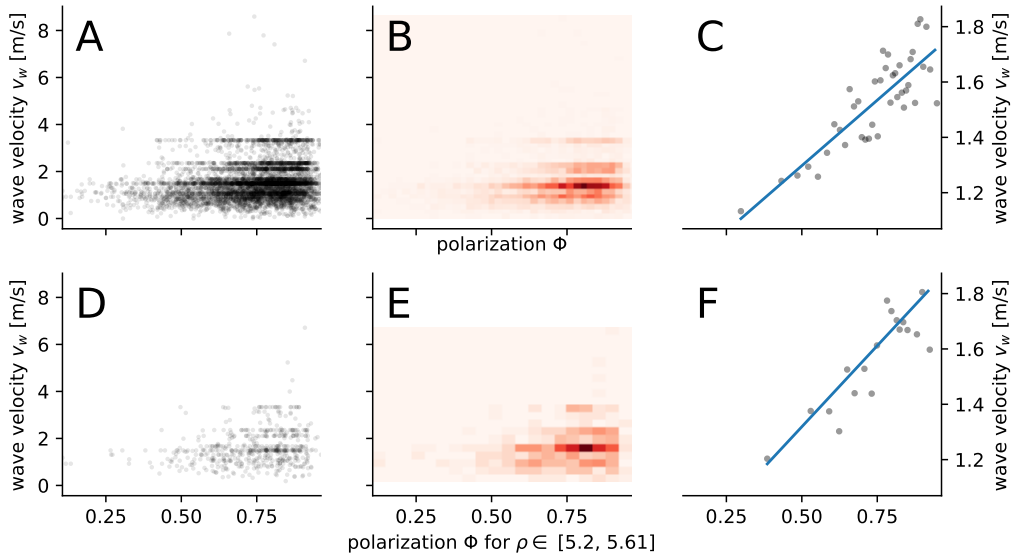

Figure C.4: Dependence of wave velocity v_w on the **polarization** of the full data set (**A, B, C**) and on polarizations corresponding to boxes with a density $\rho \in [5.2, 5.61]$ (**D, E, F**). **A, E**: raw data points. **B, F**: 2D histogram of the raw data. **C, G**: accumulated data with each data points representing the same number of raw data points (equal weights) and a linear fit (blue line).

Bibliography

Alós, J., Palmer, M., and Arlinghaus, R. (2012). Consistent Selection towards Low Activity Phenotypes When Catchability Depends on Encounters among Human Predators and Fish. *PLoS One*, 7(10):e48030.

Andersen, K. H., Marty, L., and Arlinghaus, R. (2018). Evolution of boldness and life history in response to selective harvesting. *Can. J. Fish. Aquat. Sci.*, 75(2):271–281.

Angelani, L. (2012). Collective Predation and Escape Strategies. *Phys. Rev. Lett.*, 109(11):118104.

Attanasi, A., Cavagna, A., Del Castello, L., Giardina, I., Melillo, S., Parisi, L., Pohl, O., Rossaro, B., Shen, E., Silvestri, E., and Viale, M. (2014a). Collective Behaviour without Collective Order in Wild Swarms of Midges. *PLoS Comput. Biol.*, 10(7):e1003697.

Attanasi, A., Cavagna, A., Del Castello, L., Giardina, I., Melillo, S., Parisi, L., Pohl, O., Rossaro, B., Shen, E., Silvestri, E., and Viale, M. (2014b). Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms. *Phys. Rev. Lett.*, 113(23):238102.

Axelsen, B. E., Anker-Nilssen, T., Fossum, P., Kvamme, C., and Nøttestad, L. (2001). Pretty patterns but a simple strategy: predator-prey interactions between juvenile herring and Atlantic puffins observed with multibeam sonar. *Can. J. Zool.*, 79(9):1586–1596.

Bainbridge, R. (1958). The speed of swimming of fish as related to size and to the frequency and the amplitude of the tail beat. *J. Exp. Bio.*, 35(1):109–133.

Bak, P., Chen, K., and Creutz, M. (1989). Self-organized criticality in the 'Game of Life'. *Nature*, 342:780–782.

Bak, P. and Sneppen, K. (1993). Punctuated Equilibribum and Criticality in a simple model of evolution. *Phys. Rev. Lett.*, 71(24):4083–4086.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., and Zdravkovic, V. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proc. Natl. Acad. Sci.*, 105(4):1232–1237.

Barnett, L., Brown, J., and Bossomaier, T. (2017). Anomalous behaviour of mutual information in finite flocks. *EPL (Europhysics Lett.*, 120(3):38005.

Bastien, R. and Romanczuk, P. (2020). A model of collective behavior based purely on vision. *Sci. Adv.*, 6(6):1–10.

Beauchamp, G. (2003). Group-size effects on vigilance: A search for mechanisms. *Behav. Process.*, 63(3):111–121.

Beauchamp, G. (2008). What is the magnitude of the group-size effect on vigilance? *Behav. Ecol.*, 19(6):1361–1368.

Beauchamp, G. and Ruxton, G. D. (2008). Disentangling risk dilution and collective detection in the antipredator vigilance of semipalmated sandpipers in flocks. *Anim. Behav.*, 75(6):1837–1842.

Beggs, J. M. and Timme, N. (2012). Being critical of criticality in the brain. *Front. Physiol.*, 3 JUN(June):1–14.

Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J., and Couzin, I. D. (2013). Emergent Sensing of Complex Environments by Mobile Animal Groups. *Science (80-.).*, 339(6119):574–576.

Bialek, W., Cavagna, A., Giardina, I., Mora, T., Pohl, O., Silvestri, E., Viale, M., and Walczak, A. M. (2014). Social interactions dominate speed control in poising natural flocks near criticality. *Proc. Natl. Acad. Sci. U. S. A.*, 111(20):7212–7217.

Bierbach, D., Arias-Rodriguez, L., and Plath, M. (2018). Intrasexual competition enhances reproductive isolation between locally adapted populations. *Curr. Zool.*, 64(1):125–133.

Binder, K. and Heermann, D. W. (2019). *Monte Carlo Simulation in Statistical Physics*. Graduate Texts in Physics. Springer International Publishing, Cham.

Bode, N. W. F., Faria, J. J., Franks, D. W., Krause, J., and Wood, A. J. (2010). How perceived threat increases synchronization in collectively moving animal groups. *Proc. R. Soc. B Biol. Sci.*, 277(1697):3065–3070.

Brush, E. R., Leonard, N. E., and Levin, S. A. (2016). The content and availability of information affects the evolution of social-information gathering strategies. *Theor. Ecol.*, 9(4):455–476.

Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., and Simpson, S. J. (2006). From Disorder to Order in Marching Locusts. *Science (80-.).*, 312(5778):1402–1406.

Bumann, D., Rubenstein, D., and Krause, J. (1997). Mortality Risk of Spatial Positions in Animal Groups: the Danger of Being in the Front. *Behaviour*, 134(13-14):1063–1076.

Calovi, D. S., Litchinko, A., Lecheval, V., Lopez, U., Pérez Escudero, A., Chaté, H., Sire, C., and Theraulaz, G. (2018). Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. *PLOS Comput. Biol.*, 14(1):e1005933.

Calovi, D. S., Lopez, U., Schuhmacher, P., Chate, H., Sire, C., and Theraulaz, G. (2015). Collective response to perturbations in a data-driven fish school model. *J. R. Soc. Interface*, 12(104):20141362–20141362.

Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., and Viale, M. (2010). Scale-free correlations in starling flocks. *Proc. Natl. Acad. Sci.*, 107(26):11865–11870.

Cavagna, A., Del Castello, L., Giardina, I., Grigera, T., Jelic, A., Melillo, S., Mora, T., Parisi, L., Silvestri, E., Viale, M., and Walczak, A. M. (2014). Flocking and Turning: a New Model for Self-organized Collective Motion. *J. Stat. Phys.*, 158(3):601–627.

Cavagna, A., Giardina, I., and Grigera, T. S. (2018). The physics of flocking: Correlation as a compass from experiments to theory. *Phys. Rep.*, 728:1–62.

Chaté, H., Ginelli, F., Grégoire, G., and Raynaud, F. (2008). Collective motion of self-propelled particles interacting without cohesion. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 77(4):1–15.

Chicoli, A. and Paley, D. A. (2016). Probabilistic information transmission in a network of coupled oscillators reveals speed-accuracy trade-off in responding to threats. *Chaos An Interdiscip. J. Nonlinear Sci.*, 26(11):116311.

Christensen, K. and Moloney, N. R. (2005). *Complexity and Criticality*, volume 1 of *Imperial College Press Advanced Physics Texts*. Published by Imperial College Press and distributed by World Scientific Publishing Co.

Couzin, I. D., Ioannou, C. C., Demirel, G., Gross, T., Torney, C. J., Hartnett, A., Conradt, L., Levin, S. A., and Leonard, N. E. (2011). Uninformed Individuals Promote Democratic Consensus in Animal Groups. *Science (80-.).*, 334(6062):1578–1580.

Couzin, I. D., Krause, J., Franks, N. R., and Levin, S. A. (2005). Effective leadership and decisionmaking in animal groups on the move. *Nature*, 13(4):513–516.

Couzin, I. D., Krause, J., James, R., Ruxton, G. D., and Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1):1–11.

Culumber, Z. W., Hopper, G. W., Barts, N., Passow, C. N., Morgan, S., Brown, A., Arias-Rodriguez, L., and Tobler, M. (2016). Habitat use by two extremophile, highly endemic, and critically endangered fish species (Gambusia eurystoma and Poecilia sulphuraria ; Poeciliidae). *Aquat. Conserv. Mar. Freshw. Ecosyst.*, 26(6):1155–1167.

D'Agostino, D., Jimenez, C., Reader, T., Hadjioannou, L., Heyworth, S., Aplikioti, M., Argyrou, M., and Feary, D. (2020). Behavioural traits and feeding ecology of Mediterranean lionfish and naiveté of native species to lionfish predation. *Mar. Ecol. Prog. Ser.*, 638:123–135.

Daniels, B. C., Krakauer, D. C., and Flack, J. C. (2017). Control of finite critical behaviour in a small-scale social system. *Nat. Commun.*, 8(1):14301.

Demšar, J., Hemelrijk, C. K., Hildenbrandt, H., and Bajec, I. L. (2015). Simulating predator attacks on schools: Evolving composite tactics. *Ecol. Modell.*, 304:22–33.

Dodds, P. S. and Watts, D. J. (2004). Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.*, 92(21):1–4.

Domenici, P. (2001). The scaling of locomotor performance in predator–prey encounters: from fish to killer whales. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.*, 131(1):169–182.

Domenici, P. and Batty, R. S. (1997). Escape behaviour of solitary herring (Clupea harengus) and comparisons with schooling individuals. *Mar. Biol.*, 128(1):29–38.

Doran, C., Bierbach, D., Klamser, P., Lukas, J., Landgraf, T., Klenz, H., Habedank, M., Rodriguez, L. A., Romanczuk, P., and Krause, J. (2020). Collective Fish Waves Delay Bird Attacks. *Prep.*

Edwards, A. M., Best, E. C., Blomberg, S. P., and Goldizen, A. W. (2013). Individual traits influence vigilance in wild female eastern grey kangaroos. *Aust. J. Zool.*, 61(4):332.

FAO (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals.* FAO, Rome.

Feinerman, O., Pinkoviezky, I., Gelblum, A., Fonio, E., and Gov, N. S. (2018). The physics of cooperative transport in groups of ants. *Nat. Phys.*, 14(7):1–11.

Francisco, F. A., Nührenberg, P., and Jordan, A. (2020). High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems. *Mov. Ecol.*, 8(1):2020.02.25.963926.

Gardiner, C. W. (2004). *Handbook of stochastic methods.* Number ARRAY(0x55e870d821c0) in Springer series in synergetics. Springer, Berlin ; Heidelberg [u.a.], 3. ed. edition.

Gautrais, J., Ginelli, F., Fournier, R., Blanco, S., Soria, M., Chaté, H., and Theraulaz, G. (2012). Deciphering Interactions in Moving Animal Groups. *PLoS Comput. Biol.*, 8(9):e1002678.

Gelblum, A., Pinkoviezky, I., Fonio, E., Ghosh, A., Gov, N., and Feinerman, O. (2015). Ant groups optimally amplify the effect of transiently informed individuals. *Nat. Commun.*, 6.

Gerlotto, F., Bertrand, S., Bez, N., and Gutierrez, M. (2006). Waves of agitation inside anchovy schools observed with multibeam sonar: a way to transmit information in response to predation. *ICES J. Mar. Sci.*, 63(8):1405–1417.

Ginelli, F. and Chaté, H. (2010). Relevance of Metric-Free Interactions in Flocking Phenomena. *Phys. Rev. Lett.*, 105(16):168103.

Godin, J.-G., Classon, L., and Abrahams, M. (1988). Group Vigilance and Shoal Size in a Small Characin Fish. *Behaviour*, 104(1-2):29–40.

Godin, J. G. J. and Morgan, M. J. (1985). Predator avoidance and school size in a cyprinodontid fish, the banded killifish (Fundulus diaphanus Lesueur). *Behav. Ecol. Sociobiol.*, 16(2):105–110.

Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., and Couzin, I. D. (2019). Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife*, 8:1–42.

Großmann, R., Schimansky-Geier, L., and Romanczuk, P. (2012). Active Brownian particles with velocity-alignment and active fluctuations. *New J. Phys.*, 14.

Guerra, A. S., Kao, A. B., McCauley, D. J., and Berdahl, A. M. (2020). Fisheries-induced selection against schooling behaviour in marine fishes. *Proc. R. Soc. B Biol. Sci.*, 287(1935):20201752.

Guttal, V. and Couzin, I. D. (2010). Social interactions, information use, and the evolution of collective migration. *Proc. Natl. Acad. Sci.*, 107(37):16172–16177.

Haldeman, C. and Beggs, J. M. (2005). Critical branching captures activity in living neural networks and maximizes the number of metastable states. *Phys. Rev. Lett.*, 94(5):1–4.

Hamilton, W. D. (1971). Geometry for the selfish herd. *J. Theor. Biol.*, 31(2):295–311.

Handegard, N. O., Boswell, K. M., Ioannou, C. C., Leblanc, S. P., Tjøstheim, D. B., and Couzin, I. D. (2012). The Dynamics of Coordinated Group Hunting and Collective Information Transfer among Schooling Prey. *Curr. Biol.*, 22(13):1213–1217.

Hansen, M. J., Krause, S., Breuker, M., Kurvers, R. H. J. M., Dhellemmes, F., Viblanc, P. E., Müller, J., Mahlow, C., Boswell, K., Marras, S., Domenici, P., Wilson, A. D. M., Herbert-Read, J. E., Steffensen, J. F., Fritsch, G., Hildebrandt, T. B., Zaslansky, P., Bach, P., Sabarros, P. S., and Krause, J. (2020). Linking hunting weaponry to attack strategies in sailfish and striped marlin. *Proc. R. Soc. B Biol. Sci.*, 287(1918).

Hansen, N., Akimoto, Y., and Baudis, P. (2019). CMA-ES/pycma on Github.

Hansen, N. and Ostermeier, A. (1996). Nikolaus Hansen & Andreas Ostermeier. *Gener. J. Am. Soc. Aging*, pages 312–317.

Harpaz, R., Tkačik, G., and Schneidman, E. (2017). Discrete modes of social information processing predict individual behavior of fish in a group. *Proc. Natl. Acad. Sci.*, 114(38):10149–10154.

Hein, A. M., Rosenthal, S. B., Hagstrom, G. I., Berdahl, A., Torney, C. J., and Couzin, I. D. (2015). The evolution of distributed sensing and collective computation in animal populations. *Elife*, 4(DECEMBER2015):1–43.

Heino, M., Díaz Pauli, B., and Dieckmann, U. (2015). Fisheries-Induced Evolution. *Annu. Rev. Ecol. Evol. Syst.*, 46(1):461–480.

Hemelrijk, C. K. and Kunz, H. (2005). Density distribution and size sorting in fish schools: An individual-based model. *Behav. Ecol.*, 16(1):178–187.

Herbert-Read, J. E., Buhl, J., Hu, F., Ward, A. J. W., and Sumpter, D. J. T. (2015). Initiation and spread of escape waves within animal groups. *R. Soc. Open Sci.*, 2(4):140355.

Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J., and Ward, A. J. (2011). Inferring the rules of interaction of shoaling fish. *Proc. Natl. Acad. Sci. U. S. A.*, 108(46):18726–18731.

Herbert-Read, J. E., Rosén, E., Szorkovszky, A., Ioannou, C. C., Rogell, B., Perna, A., Ramnarine, I. W., Kotrschal, A., Kolm, N., Krause, J., and Sumpter, D. J. T. (2017). How predation shapes the social interaction rules of shoaling fish. *Proc. R. Soc. B Biol. Sci.*, 284(1861):20171126.

Hess, S., Fischer, S., and Taborsky, B. (2016). Territorial aggression reduces vigilance but increases aggression towards predators in a cooperatively breeding fish. *Anim. Behav.*, 113:229–235.

Hidalgo, J., Grilli, J., Suweis, S., Munoz, M. A., Banavar, J. R., and Maritan, A. (2014). Information-based fitness and the emergence of criticality in living systems. *Proc. Natl. Acad. Sci.*, 111(28):10095–10100.

Howland, H. C. (1974). Optimal strategies for predator avoidance: the relative importance of speed and manoeuvrability. *J. Theor. Biol.*, 47(2):333–350.

James, R., Bennett, P. G., and Krause, J. (2004). Geometry for mutualistic and selfish herds: the limited domain of danger. *J. Theor. Biol.*, 228(1):107–113.

Jamie Wood, A. (2010). Strategy selection under predation; evolutionary analysis of the emergence of cohesive aggregations. *J. Theor. Biol.*, 264(4):1102–1110.

Janosov, M., Virágh, C., Vásárhelyi, G., and Vicsek, T. (2017). Group chasing tactics: how to catch a faster prey. *New J. Phys.*, 19(5):053003.

Jhawar, J., Morris, R. G., Amith-Kumar, U. R., Danny Raj, M., Rogers, T., Rajendran, H., and Guttal, V. (2020). Noise-induced schooling of fish. *Nat. Phys.*, 16(4):488–493.

Jolles, J. W., Boogert, N. J., Sridhar, V. H., Couzin, I. D., and Manica, A. (2017). Consistent Individual Differences Drive Collective Behavior and Group Functioning of Schooling Fish. *Curr. Biol.*, 27(18):2862–2868.e7.

Karvonen, A., Rintamäki, P., Jokela, J., and Valtonen, E. T. (2010). Increasing water temperature and disease risks in aquatic systems: Climate change increases the risk of some, but not all, diseases. *Int. J. Parasitol.*, 40(13):1483–1488.

Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C., and Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl. Acad. Sci. U. S. A.*, 108(46):18720–18725.

Khaluf, Y., Ferrante, E., Simoens, P., and Huepe, C. (2017). Scale invariance in natural and artificial collective systems: a review. *J. R. Soc. Interface*, 14(136):20170662.

Kingsland, S. E. (1985). *Modeling Nature: Episodes in the History of Population Ecology*. University of Chicago Press.

Klamser, P. P. and Romanczuk, P. (2020). Collective predator evasion: Putting the criticality hypothesis to the test. *Rev. PLOS Comput. Biol.*

Kleinhappel, T. K., Pike, T. W., and Burman, O. H. P. (2019). Stress-induced changes in group behaviour. *Sci. Rep.*, 9(1):1–9.

Kolpas, A., Busch, M., Li, H., Couzin, I. D., Petzold, L., and Moehlis, J. (2013). How the Spatial Position of Individuals Affects Their Influence on Swarms: A Numerical Comparison of Two Popular Swarm Dynamics Models. *PLoS One*, 8(3):e58525.

Krause, J. (1994). Differential fitness returns in relation to spatial position in groups. *Biol. Rev.*, 69(2):187–206.

Krause, J. and Ruxton, G. D. (2002). *Living in Groups*. Oxford Univ. Press, Oxford.

Kyriakopoulos, N., Chaté, H., and Ginelli, F. (2019). Clustering and anisotropic correlated percolation in polar flocks. *Phys. Rev. E*, 100(2):022606.

Landeau, L. and Terborgh, J. (1986). Oddity and the 'confusion effect' in predation. *Anim. Behav.*, 34(5):1372–1380.

Langton, C. G. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Phys. D*, 42:12–37.

Lei, L., Escobedo, R., Sire, C., and Theraulaz, G. (2020). Computational and robotic modeling reveal parsimonious combinations of interactions between individuals in schooling fish. *PLOS Comput. Biol.*, 16(3):e1007194.

Lemasson, B. H., Anderson, J. J., and Goodwin, R. A. (2009). Collective motion in animal groups from a neurobiological perspective: The adaptive benefits of dynamic sensory loads and selective attention. *J. Theor. Biol.*, 261(4):501–510.

Lemasson, B. H., Anderson, J. J., and Goodwin, R. A. (2013). Motion-guided attention promotes adaptive communications during social navigation. *Proc. Biol. Sci.*, 280(1754):20122003.

Lorenzen, K. (2000). Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. *Can. J. Fish. Aquat. Sci.*, 57(12):2374–2381.

Lukas, J. (2018). private communication.

Lukas, J., Romanczuk, P., Klenz, H., Klamser, P., Rodriguez, L. A., Krause, J., and Bierbach, D. (2020). Acoustic and visual stimuli combined promote stronger responses to aerial predation in fish. *Rev. Behav. Ecol.*

MacGregor, H. E. A., Herbert-Read, J. E., and Ioannou, C. C. (2020). Information can explain the dynamics of group order in animal collective behaviour. *Nat. Commun.*, 11(1):2737.

Marconi, U. M. B., Puglisi, A., Rondoni, L., and Vulpiani, A. (2008). Fluctuation-dissipation: Response theory in statistical physics. *Phys. Rep.*, 461(4-6):111–195.

Mateo, D., Kuan, Y. K., and Bouffanais, R. (2017). Effect of Correlations in Swarms on Collective Response. *Sci. Rep.*, 7(1):10388.

Meron, E. (1992). Pattern formation in excitable media. *Phys. Rep.*, 218(1):1–66.

Milinski, M. (2010a). Do all Members of a Swarm Suffer the same Predation? *Z. Tierpsychol.*, 45(4):373–388.

Milinski, M. (2010b). Experiments on the Selection by Predators against spatial Oddity of their Prey. *Z. Tierpsychol.*, 43(3):311–325.

Monk, C. T., Barbier, M., Romanczuk, P., Watson, J. R., Alós, J., Nakayama, S., Rubenstein, D. I., Levin, S. A., and Arlinghaus, R. (2018). How ecology shapes exploitation: a framework to predict the behavioural response of human and animal foragers along exploration-exploitation trade-offs. *Ecol. Lett.*, 21(6):779–793.

Mora, T. and Bialek, W. (2011). Are Biological Systems Poised at Criticality? *J. Stat. Phys.*, 144(2):268–302.

Muñoz, M. A. (2018). Colloquium: Criticality and dynamical scaling in living systems. *Rev. Mod. Phys.*, 90(3):31001.

Nagy, M., Ákos, Z., Biro, D., and Vicsek, T. (2010). Hierarchical group dynamics in pigeon flocks. *Nature*, 464(7290):890–893.

Newman, M. E. J. (2002). Mixing patterns in networks. *Phys. Rev. E*, 67(2):26126.

Nissanov, J., Eaton, R. C., and DiDomenico, R. (1990). The motor output of the Mauthner cell, a reticulospinal command neuron. *Brain Res.*, 517(1-2):88–98.

Nykter, M., Price, N. D., Larjo, A., Aho, T., Kauffman, S. A., Yli-Harja, O., and Shmulevich, I. (2008). Critical networks exhibit maximal information diversity in structure-dynamics relationships. *Phys. Rev. Lett.*, 100(5):1–4.

Olson, R. S., Hintze, A., Dyer, F. C., Knoester, D. B., and Adami, C. (2013). Predator confusion is sufficient to evolve swarming behaviour. *J. R. Soc. Interface*, 10(85):20130305.

Olson, R. S., Knoester, D. B., and Adami, C. (2016). Evolution of Swarming Behavior Is Shaped by How Predators Attack. *Artif. Life*, 22(3):299–318.

Oufiero, C. E. and Garland, T. (2009). Repeatability and correlation of swimming performances and size over varying time-scales in the guppy (Poecilia reticulata). *Funct. Ecol.*, 23(5):969–978.

Packard, N. H. (1988). Adaptation Toward the Edge of Chaos. In Kelso, J., Mandell, A., and Shlesinger, M., editors, *Dyn. Patterns Complex Syst.* Singapore, World Scientific.

Pérez-Escudero, A., Vicente-Page, J., Hinz, R. C., Arganda, S., and de Polavieja, G. G. (2014). idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods*, 11(7):743–748.

Pita, D., Moore, B. A., Tyrrell, L. P., and Fernández-Juricic, E. (2015). Vision in two cyprinid fish: implications for collective behavior. *PeerJ*, 3(8):e1113.

Pitcher, T. (1996). Adaptive behaviour of herring schools in the Norwegian Sea as revealed by high-resolution sonar. *ICES J. Mar. Sci.*, 53(2):449–452.

Price, V. E., Auster, P. J., and Kracker, L. (2013). Use of High-Resolution DIDSON Sonar to Quantify Attributes of Predation at Ecologically Relevant Space and Time Scales. *Mar. Technol. Soc. J.*, 47(1):33–46.

Quinn, T. P., Hodgson, S., Flynn, L., Hilborn, R., and Rogers, D. E. (2007). Directional selection by fisheries and the timing of sockeye salmon (Oncorhynchus nerka) migrations. *Ecol. Appl.*, 17(3):731–739.

Rahmani, P., Peruani, F., and Romanczuk, P. (2020). Flocking in complex environments—Attention trade-offs in collective information processing. *PLOS Comput. Biol.*, 16(4):e1007697.

Riofrío-Lazo, M., Arreguín-Sánchez, F., and Páez-Rosas, D. (2017). Population Abundance of the Endangered Galapagos Sea Lion Zalophus wollebaeki in the Southeastern Galapagos Archipelago. *PLoS One*, 12(1):e0168829.

Roli, A., Villani, M., Filisetti, A., and Serra, R. (2018). Dynamical Criticality: Overview and Open Questions. *J. Syst. Sci. Complex.*, 31(3):647–663.

Romanczuk, P., Couzin, I. D., and Schimansky-Geier, L. (2009). Collective Motion due to Individual Escape and Pursuit Response. *Phys. Rev. Lett.*, 102(1):010602.

Romanczuk, P. and Schimansky-Geier, L. (2011). Brownian Motion with Active Fluctuations. *Phys. Rev. Lett.*, 106(23):230601.

Rosenthal, S. B., Twomey, C. R., Hartnett, A. T., Wu, H. S., and Couzin, I. D. (2015). Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proc. Natl. Acad. Sci. U. S. A.*, 112(15):4690–4695.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J. (2019a). Inferring causation from time series in Earth system sciences. *Nat. Commun.*, 10(1):2553.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019b). Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.*, 5(11):eaau4996.

Salomon, M. and Holm-Müller, K. (2013). Towards a sustainable fisheries policy in Europe. *Fish Fish.*, 14(4):625–638.

Sarracino, A. and Vulpiani, A. (2019). On the fluctuation-dissipation relation in non-equilibrium and non-Hamiltonian systems. *Chaos An Interdiscip. J. Nonlinear Sci.*, 29(8):083132.

Sbragaglia, V., Klamser, P. P., Romanczuk, P., and Arlinghaus, R. (2020). Unexpected harvesting-induced evolution of collective behavior in a fish. *Rev. Am. Nat.*

Sbragaglia, V., López-Olmeda, J. F., Frigato, E., Bertolucci, C., and Arlinghaus, R. (2019). Fisheries-induced evolution of the circadian system and collective personality traits. *bioRxiv*, page 622043.

Sbragaglia, V., Morroni, L., Bramanti, L., Weitzmann, B., Arlinghaus, R., and Azzurro, E. (2018). Spearfishing modulates flight initiation distance of fishes: the effects of protection, individual size, and bearing a speargun. *ICES J. Mar. Sci.*, 75(5):1779–1789.

Schreiber, T. (2000). Measuring Information Transfer. *Phys. Rev. Lett.*, 85(2):461–464.

Similä, T. (1997). Sonar observations of killer whales (Orcinus orca) feeding on herring schools. *Aquat. Mamm.*, 23(3):119–126.

Sogard, S. M. (1997). Size-selective mortality in the juvenile stage of teleost fishes: A review. *Bull. Mar. Sci.*, 60(3):1129–1157.

Sosna, M. M. G., Twomey, C. R., Bak-Coleman, J., Poel, W., Daniels, B. C., Romanczuk, P., and Couzin, I. D. (2019). Individual and collective encoding of risk in animal groups. *Proc. Natl. Acad. Sci. U. S. A.*, 116(41):20556–20561.

Speedie, N. and Gerlai, R. (2008). Alarm substance induced behavioral responses in zebrafish (Danio rerio). *Behav. Brain Res.*, 188(1):168–177.

Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., Summerhayes, C. P., Barnosky, A. D., Cornell, S. E., Crucifix, M., Donges, J. F., Fetzer, I., Lade, S. J., Scheffer, M., Winkelmann, R., and Schellnhuber, H. J. (2018). Trajectories of the Earth System in the Anthropocene. *Proc. Natl. Acad. Sci.*, 115(33):8252–8259.

Storn, R. and Price, K. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J. Glob. Optim.*, 11:341–359.

Strandburg-Peshkin, A., Twomey, C. R., Bode, N. W. F. F., Kao, A. B., Katz, Y., Ioannou, C. C., Rosenthal, S. B., Torney, C. J., Wu, H. S., Levin, S. A., and Couzin, I. D. (2013). Visual sensory networks and effective information transfer in animal groups. *Curr. Biol.*, 23(17):R709–R711.

Theodorakis, C. W. (1989). Size segregation and the effects of oddity on predation risk in minnow schools. *Anim. Behav.*, 38(3):496–502.

Torney, C. J., Lorenzi, T., Couzin, I. D., and Levin, S. A. (2015). Social information use and the evolution of unresponsiveness in collective systems. *J. R. Soc. Interface*, 12(103):20140893.

van der Post, D. J., Verbrugge, R., and Hemelrijk, C. K. (2015). The Evolution of Different Forms of Sociality: Behavioral Mechanisms and Eco-Evolutionary Feedback. *PLoS One*, 10(1):e0117027.

Vanni, F., Luković, M., and Grigolini, P. (2011). Criticality and Transmission of Information in a Swarm of Cooperative Units. *Phys. Rev. Lett.*, 107(7):078103.

Vicsek, T. T., Czirk, A., Ben-Jacob, E., Cohen, I., Shochet, O., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O. (1995). Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys. Rev. Lett.*, 75(6):1226–1229.

Ward, A. J. W., Herbert-Read, J. E., Sumpter, D. J. T., and Krause, J. (2011). Fast and accurate decisions through collective vigilance in fish shoals. *Proc. Natl. Acad. Sci.*, 108(6):2312–2315.

Waxman, D. and Gavrilets, S. (2005). 20 Questions on Adaptive Dynamics. *J. Evol. Biol.*, 18(5):1139–1154.

Wilson, D. S. (1975). A theory of group selection. *Proc. Natl. Acad. Sci. U. S. A.*, 72(1):143–146.

Wilson, D. S. (1997). Altruism And Organism: Disentangling The Themes Of Multilevel Selection Theory. *Am. Nat.*, 150(S1):S122–S134.

Wood, A. J. and Ackland, G. J. (2007). Evolving the selfish herd: emergence of distinct aggregating strategies in an individual-based model. *Proc. R. Soc. B Biol. Sci.*, 274(1618):1637–1642.

Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for global optimization: The GenSA package. *R J.*, 5(1):13–28.

Zienkiewicz, A. K., Ladu, F., Barton, D. A., Porfiri, M., and Bernardo, M. D. (2018). Data-driven modelling of social forces and collective behaviour in zebrafish. *J. Theor. Biol.*, 443:39–51.

Statutory Declaration

I, Pascal Peter Klamser, declare on oath that I completed this work on my own and that information which has been directly or indirectly taken from other sources has been noted as such. Neither this, nor a similar work, has been published or presented to an examination committee.

Berlin, July 21, 2021